

Judgments of information structure in L2 French: Nativelike performance and the Critical Period Hypothesis¹

ROBERT V. REICHLE

Abstract

Previous studies using judgments of morphosyntactic errors have shown mixed evidence for a critical period for L2 acquisition (e.g., Birdsong & Molis 2001, Johnson & Newport 1989). This study uses anomalies in the domain of information structure, the interface between syntactic form and pragmatic function, to shed light on the effect of age of arrival on L2 performance. In two experiments, high-proficiency L2 speakers of French were presented with sentence pairs containing either expected or anomalous information structure. Subjects judged each exchange as acceptable or unacceptable. A weak post-maturational effect of age on acceptability task performance was observed, along with a high degree of nativelike performance. These results are incompatible with the traditional notion of a critical period for second language acquisition in this domain of language, and also suggest that long periods of immersion in the L2 environment can lead to nativelike performance on tasks relating to information structure.

1. Introduction

Information structure (IS) can be described as the interface between syntactic form and pragmatic function, or in other words, the way in which a speaker uses cues from sentence structure to guide a hearer toward knowing what is more or less important in a sentence. These relations between the more- and less-prominent take the form of topic, focus, presupposition and assertion relations between the constituents of an utterance. The acquisition of information

1. An earlier version of this paper reporting the results of Experiment 1 appeared in the proceedings of the 19th International Conference on Foreign/Second Language Acquisition (Reichle 2009).

structure by adult L2 learners has previously been examined as it relates to the syntactic constructions employed by L2 learners in discourse and narrative (e.g., Carroll et al. 2000). Some previous studies have used production data to examine the acquisition of finiteness in L2 learners of French (Schlyter 2003) and German (Dimroth et al. 2003; Dimroth 2008); others have studied anaphora in narrative (Carroll and Lambert 2003), the production of topic markers (Ferdinand 2002), and the grammatical means used to organize information in learner varieties of German and English (Carroll & von Stutterheim 2003). While many existing studies have examined the assignment of information structure to sentences, fewer studies have looked at L2 learners' ability to identify and process anomalies in the information structure of a sentence, and those that have looked at this have targeted L2 learners' sensitivity to information structure distinctions communicated by word order (e.g., Wilson et al. 2007 for German; Kaiser & Trueswell 2004 for Finnish).

The nature of information structure includes elements of semantics (as it deals with the relations between real-world referents, and how those referents are highlighted in speech) and syntax (since different word orders or syntactic constructions are tied to different IS configurations). French information structure is an interesting case because, unlike English IS (which relies heavily on prosody and word accent to distinguish between topic and focus elements), French IS makes extensive use of syntactic constructions to express information structural distinctions.

Previous research on second language acquisition has shown effects for age of arrival in the L2 environment when L2 subjects engage in grammaticality judgment tasks; however, tasks relating to violations of expected information structure have so far been underrepresented in the body of research (Wilson et al. 2007 and Kaiser & Trueswell 2004 are notable exceptions). The present study uses behavioral data from two experiments to address the following research questions: Are L2 learners capable of performing like native speakers when making judgments about the acceptability of information structure constructions? Among L2 subjects, are there effects of age of arrival (AoA) on the ability to judge information structure anomalies? If so, does the geometry of these effects support a critical period for this facet of second language acquisition?

2. Information structure in French

Linguistic information is structured by three parameters on the exchange between speaker and hearer: (1) the speaker's assumptions concerning the hearer's knowledge or belief state, (2) the hearer's temporary state of consciousness, and (3) the hearer's state of interest in a given topic of conversation

(Lambrecht et al. 2006). The present study is mainly concerned with these parameters insofar as they relate to the notions of topic and focus.

Parameter 1 – the hearer’s knowledge or belief state – is relevant to the speaker’s pragmatic presuppositions (the propositions and referents that are assumed to be already known to the interlocutor) and pragmatic assertions (the propositions and referents that are assumed by the speaker to be known only as a result of the interlocutor’s comprehending the utterance). The speaker takes into account these beliefs about the hearer’s state of knowledge and, in conjunction with his beliefs regarding the hearer’s state of interest (i.e., Parameter 3), the speaker accords a level of topic or focus to the constituents within the utterance. Focus can be defined as the “element of information whereby the assertion differs from the presupposition” and topic as the constituent “relative to which a predication is to be assessed as relevant information” (Lambrecht et al. 2006; see also Lambrecht 1994). In other words, the focal element is granted a level of prominence greater than that of the rest of the proposition, and it is telling the hearer something new or different with regards to the topic, which is what the sentence is “about”.

Cross-linguistically, languages use a variety of methods to allow for the marking of a constituent as focal. French has several constructions at its disposal for this purpose, most of which rely on clefting, dislocation or some other variation from canonical word order to indicate different types of focal relations (Lambrecht 1994). An example of these constructions that is relevant to the present study is the *c’est* cleft. A cleft construction expresses a proposition by way of biclausal syntax; in French, this typically consists of a copula clause and a relative clause that share a co-indexed referent (Lambrecht 2001). Compare the following sentences, which both express the proposition of a piece of pizza being located on a table (throughout, word accent in English glosses is represented by italics):

- (1) a. *La pizza est sur la table.*
 the pizza is on the table
 ‘The pizza is on the table.’
 b. *C’est la pizza qui est sur la table.*
 it is the pizza that is on the table
 ‘The *pizza* is on the table.’

The two sentences identify the same spatial relationship between the referents of *la pizza* and *la table*, but they represent different configurations of IS parameters. (1a) exemplifies the canonical written French sentence, and as such illustrates the unmarked IS configuration, where *la pizza* (the grammatical subject) functions as the topic constituent, and the predicate *N est sur la table* is the focal element. Sentence (1b), on the other hand, is an argument-focus sentence, in which the cleft construction puts the argument *la pizza* in focus, while

the proposition *N est sur la table* is the topic constituent. For a sentence like (1b) to be uttered felicitously, there must be a previously existing context asserting the presence of some object on the table. Once the proposition *N est sur la table* has been asserted, its topicality is now presupposed, allowing the speaker of sentence (1b) to place focus on the referent that is located on the table, rather than placing focus on the fact that there is some object on the table as in sentence (1a). Consider the exchange in (2):

- (2) X: *C'est quoi qui se trouve sur la table?*
 it is what that is located on the table
 'What is located on the table?'
 Y: *C'est la pizza qui se trouve sur la table.*
 it is the pizza that is located on the table
 'The pizza is located on the table.'
 Y: #*C'est sur la table que se trouve la pizza.*
 it is on the table that the pizza is located
 #'The pizza is located on the table.'

Speaker X's turn in (2a) establishes the proposition that something is located on the table; furthermore, it places argument focus on the unknown item. Speaker Y's response in (2b) is an appropriate follow-up – Y keeps the focus on the previously unknown item (*la pizza*), and retains the proposition of something being on the table as a topical element. The response in (2c), however, is not contextually appropriate (as indicated by #). Here the speaker treats the previously unknown item (*la pizza*) as a topic, despite the fact that its identity has not yet been asserted or established as a topic. In addition, (2c) inappropriately treats *sur la table* as a focus constituent, even though (2a) established it as a topic. Note that there is nothing wrong with the referentiality or truth value of (2c) – the sentence accurately reflects the spatial relations of the referents – nor are there any syntactic violations. (2c) is anomalous strictly from an information structural/pragmatic point of view.

This ability to identify focus constituents in French by means of syntactic constructions such as the *c'est* cleft makes the language well suited for experimental studies in the written modality. By limiting the presentation of stimuli to the written modality, one can eliminate subjects' knowledge of phonetic, phonological and suprasegmental information as variables impacting their ability to process information structure. This is of particular use when looking at L2 speakers of French, as it allows us to limit the scope of inquiry to the processing of information structure anomalies.

In various studies of L2 processing, a negative correlation has been found between AoA and L2 proficiency. It has been posited that after a certain AoA, L2 acquisition becomes qualitatively different; for example, Pinker (1994) has proposed that the neural structures specialized for language acquisition are dis-

mantled in adulthood once they are no longer needed, resulting in a critical or optimal period for first and second language acquisition. Johnson & Newport (1989) advanced a similar hypothesis, positing the existence of a well-adapted language learning capability in childhood that is reduced with maturation. Such hypotheses are but two of the many formulations of the Critical Period Hypothesis for L2 acquisition. Singleton (2005) likens the Critical Period Hypothesis, with its many different definitions and end-points, to the “hydra, whose multiplicity of heads and capacity to produce new heads rendered it impossible to deal with” (Singleton 2005: 280). He suggests that, even if the various definitions and goal posts linked to different versions of the hypothesis were all reconciled and combined into one definition, the new unified version of the hypothesis would be too vague to be verifiable. Nevertheless, an inventory of the most prominent definitions of the critical period for second language acquisition is useful inasmuch as it provides guidelines by which to analyze data from L2 learners of French presented with violations of information structure.

3. Previous formulations of the Critical Period Hypothesis for L2 acquisition

Generically, a critical period can be defined as “the temporal span during which an organism displays a heightened sensitivity to certain environmental stimuli, the presence of which is required to trigger a developmental event. Typically, there is an abrupt onset or increase of sensitivity, a plateau of peak sensitivity, followed by a gradual offset or decline, with subsequent flattening of the degree of sensitivity” (Birdsong 2005: 111; see also Bornstein 1989, Colombo 1982). Under this definition, a critical period starts at the peak level of sensitivity and includes any decline and subsequent reduced level of sensitivity that follows. This representation of a critical period, containing discontinuities of linearity only between peak and offset and between offset and floor, is sometimes referred to as the “stretched Z” representation (Newport 1991; see also Birdsong 2005 for elaboration). A corollary of this is that a pattern of data showing an unbounded decline, or one *without* the requisite discontinuity shifting between a peak and an offset, would be inconsistent with this formulation of a critical period.

Lenneberg (1967) originally popularized the idea of a critical period for language acquisition ending around puberty, claiming that after this period, L1 acquisition would be impossible. Applying this to second language acquisition, Bley-Vroman (1989) took a strong stance on the discrepancy in outcomes between adult and child language acquisition, characterizing them as “fundamentally different” and claiming that adults do not have any access to the acquisition system available to children. According to a strong version of the

Critical Period Hypothesis for second language acquisition, speakers who begin to acquire an L2 after the end of the proposed critical period should exhibit a substantial deficit in production and comprehension accuracy compared to L2 speakers who acquire the language before the end of said period; the end of this period has variously been posited to coincide with the beginning or end of puberty or maturation (Birdsong 2006; Singleton & Ryan 2004). However, if the period during which cognitive and maturational changes take place has already passed, all L2 learners who begin to acquire the language after the end of maturation should be equally poor at acquiring it. Johnson & Newport (1989: 90) characterized this account by suggesting that “the age effect is present during a time of ongoing biological and cognitive maturation and absent after maturation is complete (i.e., at puberty).” In their formulation of the critical period, “there should be a consistent decline in performance over age for those exposed to the language before puberty, but no systematic relationship to age of exposure, and a leveling off of ultimate performance, among those exposed to the language after puberty” (1989: 79). Therefore, under their definition of the Critical Period Hypothesis for L2 acquisition, one would expect to see age effects between pre- and post-critical period groups, but one would not expect to see a *continuous* age effect for those who were first exposed to an L2 *after* the end of the critical period. Post-maturational age effects would be seen as evidence against the Johnson & Newport (1989) version of the Critical Period Hypothesis for L2 acquisition.

Johnson & Newport (1989) also proposed that in the presence of a critical period, the incidence of nativelikeness among late-AoA L2 learners is zero. It follows that instances of nativelike acquisition of a feature of language among late-AoA L2 learners would serve as counter-evidence to a critical period for the acquisition of that feature.

As a body of research on the putative critical period developed in the years since Lenneberg’s original formulation of the Critical Period Hypothesis, other refinements to the definition of the critical period for L2 acquisition have more narrowly defined the scope of inquiry. Singleton & Ryan (2004) provide a general definition of a biologically based critical period using three criteria:

- (1) it must apply to only a limited set of functions or behaviors (the authors cite the imprinting phenomenon of newborn ducklings as a non-linguistic example);
- (2) it must be bounded by “well-defined and predictable termini”;
- (3) outside of the period bounded by these termini, the organism cannot acquire the behavior in question.

A critical period defined by these criteria could be visually represented by what Birdsong (2006) calls a “stretched 7” – during the critical period, there is a horizontal line representing nativelike attainment, followed by a downward sloping line after the terminus of the period. A critical period defined by Johnson &

Newport's (1989) criteria, on the other hand, resembles what Birdsong (2006) calls a "stretched Z" or a "stretched L", i.e., performance declines during maturation (represented by a downward-sloping line), and this downward slope changes to a horizontal line at the end of maturation.

All of these definitions of a critical period can be described using three criteria:

- (1) the absence of nativelike acquisition among late-AoA L2 learners;
- (2) the presence of an eventual flattening in the AoA-performance relationship;
- (3) the timing of discontinuities in the age function at transition points between periods of declines and leveling off of performance.

For example, Johnson & Newport's (1989) version of a critical period would:

- (1) have no nativelike acquisition among late-AoA L2 learners;
- (2) have an eventual flattening of the AoA-performance relationship;
- (3) have discontinuities in the age function at the beginning and end of puberty.

Previous studies, such as those described below, have assessed the existence of a putative critical period in these terms. In relation to criterion (1), the presence of nativelike acquisition can be considered evidence against the Critical Period Hypothesis. In relation to criterion (2), the presence of a flattening in the AoA-performance relationship post-maturation can be considered evidence of a "stretched Z" or "stretched L" geometry and therefore evidence for a critical period as defined by Johnson & Newport (1989); the absence of this leveling off can be taken as evidence against the presence of such a critical period. In relation to criterion (3), discontinuities coinciding with the beginning, or possibly the end, of maturation could be considered evidence for a critical period, whereas discontinuities occurring after the end of maturation (or not occurring at all) should be considered evidence against the Critical Period Hypothesis.

4. Evidence for and against a critical period for L2 acquisition

Johnson & Newport (1989) conducted a study that examined native speakers of Korean and Chinese learning English as an L2. The subjects who were chosen had ages of acquisition that varied widely, and were placed into two age groups: early arrivals, who arrived in the United States before age 15, and late arrivals, who arrived after age 17. The subjects performed a grammaticality judgment task that tested several aspects of English grammar. The results showed a statistically significant relationship between AoA and performance on the task ($r = -.77, p < .01$). For members of the early AoA group, the correlation was even stronger ($r = -.87, p < .01$). For the late AoA group, there was no significant correlation ($r = -.16, p > .05$).

Johnson & Newport interpreted the group results for the late arrivals as evidence supporting a maturational model of language acquisition, since on average they had a reduced capacity for L2 acquisition compared to those who arrived as children. However, even though the averaged scores for the late AoA group appeared to show a leveling off of performance, the data for individual subjects showed what appeared to be a nearly random distribution (hence the weak correlation). Some late-arrival subjects scored high (one even scored at or near ceiling), and others scored poorly; while the group as a whole performed poorly and without a strong correlation between AoA and performance, certain individuals were exceptions to this trend. Birdsong & Molis (2001) point out the random distribution in the data for late arrivals, and contrast this with results from their own replication study (see below).

Several replications of the Johnson & Newport (1989) study have been conducted. In DeKeyser's (2000) study, native speakers of Hungarian who had immigrated to the US were given a shortened, modified version of Johnson & Newport's grammaticality judgment test of morphosyntactic constructions. The results showed a significant negative correlation between age of arrival and performance on the test for the subject group as a whole; however, no significant correlation was seen for either the early- and late-arrival groups when they were separated. The overall negative correlation was interpreted as being compatible with the results of Johnson & Newport (1989). However, the non-significant correlation for late arrivals did not indicate a flattening of performance, as predicted by the Johnson and Newport (1989) definition of a critical period. Additionally, Bialystok (2002) noted that there was a lack of discontinuity in the correlation between age of arrival and performance, as would be predicted by Johnson & Newport's definition of the critical period. In her view, these particular results did not constitute evidence supporting the Critical Period Hypothesis.

In addition to DeKeyser's (2000) study, there is a large body of work presenting evidence in favor of the Critical Period Hypothesis for second language acquisition; e.g., Johnson (1992), Johnson & Newport (1991), among others.

Not all of the studies revisiting critical period effects have found evidence supporting the claims of the Critical Period Hypothesis. Birdsong & Molis (2001) replicated the Johnson & Newport (1989) study, with several notable differences in their results. Instead of using a sample of Korean and Chinese speakers of English as an L2, this study used only native Spanish speakers, thereby eliminating the original study's confounding factor of having speakers of two typologically dissimilar languages in one group. The subject group was also larger than the one used in the original study ($n = 61$). The Spanish L1 group was divided into early ($\text{AoA} \leq 16$, $n = 29$) and late ($\text{AoA} \geq 17$, $n = 32$) arrivals into the L2 environment. The mean length of residence in the United States was 10 years. The subjects were presented with the same

recorded English sentences as in the original study and were asked to make judgments of their grammaticality.

The early arrival group scored extremely well, but there was no significant correlation between AoA and accuracy ($r = -.22, p = .22$). The distribution of the data for early arrivals is indicative of a ceiling effect – all subjects achieved perfect or near-perfect scores. This caused all data points in this group to fall roughly along a horizontal line; since there was no notable change in accuracy scores correlating with a change in AoA, one cannot properly speak of the possibility of a correlation for this group, even though the data had an orderly distribution.

The late arrival group exhibited a strong age effect ($r = -.69, p < .001$). This included data points from much older subjects (AoA = 35-45). This geometry resembled the “stretched 7” and is not compatible with Johnson & Newport’s (1989) “stretched Z” definition of a critical period. The Birdsong & Molis replication, on the other hand, did find a strong negative correlation in a data set with a more orderly distribution. This suggests that there is an age effect after the end of maturation, as opposed to a critical period after which performance is random or flat.

Birdsong & Molis (2001) also found “modest evidence of nativelike performance” among the late arrivals. The results of this study do not support a “stretched Z”-like pattern of age effects, as there is no discontinuity at the end of the putative critical period. According to the previously discussed geometric criteria, the results of this study suggest a refutation of the Critical Period Hypothesis for L2 learners.

In another study, Flege et al. (1999) carried out a series of experiments meant to test the Critical Period Hypothesis while simultaneously examining whether a putative critical period affects different domains of language equally – specifically, if it affects phonology and morphosyntax equally, and if it affects rule-based and idiosyncratic processes equally. Their control group consisted of 24 native speakers of English; the experimental group consisted of 240 native speakers of Korean who had acquired English as an L2. The Korean subjects were split into 10 subgroups based on age of arrival. There were two sections to the experiment: In the first part, subjects were aurally presented with English sentences, which they then repeated. Their output was later ranked in nativelikeness by a group of L1 speakers. In the second part of the experiment, subjects responded to a grammaticality judgment test. Most of the stimuli were taken or adapted from Johnson & Newport (1989); new stimuli were also added in order to test rule-based versus lexically based items.

Data from the phonology portion of the experiment showed that the later the age of arrival for a subject, the less nativelike his accent was rated by the native speaker judges. When comparing means, all 10 of the L2 speaker subgroups scored worse than the control group. Later ages of arrival saw decreases in

nativelikeness scores. The results from the morphosyntactic portion of the experiment differed slightly. Here, the lowest AoA group (2–6 years) did not score differently from the native controls at a level of statistical significance, but the other AoA groups did.

The geometry of the Flege et al. (1999) data does not, in general, support a critical period for second language acquisition. The CPH predicts a discontinuity in the correlation between AoA and proficiency at the onset of puberty; this was operationalized as the range of 12–15 years. In the phonology experiment, a discontinuity was present; however, it did not occur at the time of puberty, but rather occurred sometime around 5 years. The data for the morphosyntax experiment, on the other hand, was much closer to the geometry predicted by the CPH – there was a discontinuity in the 12–15 year AoA range. However, there was a correlation between AoA and proficiency for subjects whose AoA was both before *and* after the 12–15 years range; this finding is at odds with the results of Johnson & Newport (1989).

In lieu of administering grammaticality judgments tests or other psycholinguistic metrics, Bialystok, Hakuta and colleagues conducted an analysis of data from the 1990 U.S. Census (Bialystok & Hakuta 1999; Hakuta et al. 2003). They analyzed the data from 2,016,317 Spanish L1s and 324,444 Chinese L1s. The census asked respondents to rate their English ability; it also collected information about their educational background, the year of their arrival in the United States, and their age at the time of the survey. The age and year information allowed the investigators to calculate AoA for each subject. Only data from respondents with at least 10 years of residence in the United States was used. The experimenters specified that, for a discontinuity in the AoA–proficiency function to truly be indicative of the end of the putative critical period, the discontinuity would have to be independent of all factors other than maturational. For example, social changes such as a change in educational environment are indexed with age, yet are not directly due to maturation. Considering education as a factor separate from age allowed the authors to show that any potential discontinuity was independent of maturation.

The results, however, lacked a clear discontinuity at the end of the putative critical period. The experimenters proposed two possible ages for the “critical point”, i.e., the point at which a discontinuity would be observed: ages 15 and 20. An analysis of the data showed no discontinuity for either subject group at either putative critical point. Instead, proficiency steadily declined in a linear fashion.

These and other studies drawing on behavioral data provide mixed evidence for a critical period for L2 acquisition. While most of them used morphosyntactic or, in some cases, phonological performance as measures of L2 learners’ language ability, none used tasks relating to information structure. Two experiments were conducted to look at age effects on L2 subjects’ performance on

an information structure judgment task, and to determine whether these results suggest that acquisition of information structure is constrained by a critical period.

5. Experiment 1

A judgment task was developed to determine the geometry and timing of any age effects seen in the identification of information structural anomalies by learners of French as a second language. Observed post-maturational age effects would constitute evidence against a strong version of the Critical Period Hypothesis as it relates to the acquisition of information structure in the L2. Conversely, evidence of a “stretched Z” geometry of age effects would support the idea of a critical period for L2 acquisition.

5.1. Subjects

In this study, participants were asked to judge the acceptability of 60 written exchanges containing a variety of syntactic constructions used to communicate information structure in French.

Control subjects ($n = 44$) were adult native speakers of French. Subjects in the experimental group ($n = 26$) were end-state speakers of L2 French (L1 was English). Age of arrival in the target environment ranged from 1 to 34 years. All subjects were at least 18 years of age. End state was operationalized as having lived in a Francophone country, and having used French as their primary language, for four or more years. While four years is not an optimally long period of residence (Birdsong 2005), practical considerations for recruitment of participants necessitated this operationalization. The length of residency ranged from 4 to 32 years; the mean length of residency was 10 years. Nine subjects in the experimental group submitted incomplete surveys, and their data were eliminated from the analysis.

5.2. Procedure

Via e-mail, subjects received a link to an online biographic questionnaire and a 60-item survey.² Subjects were asked questions pertaining to their language

2. While the use of this type of online survey carries with it limitations on the degree of control the experimenter has over the experimental conditions, it was employed in order to increase the diversity of the subject pool at low cost; see Kraut et al. (2004) for a discussion of the limitations and benefits of this kind of survey.

use; in particular, they were asked to estimate the percentage of the time they hear and speak French at work, at home, and with friends, as well as to rate their own knowledge of French vocabulary, syntax and pronunciation on a scale of 1 to 10. Information on the subjects' language history (languages spoken, countries of residence) was also collected. L2 subjects were asked to provide information regarding their age at the time of arrival in the francophone environment and the duration of residence there.

The target stimuli consisted of constructed exchanges between two speakers, and included three different types of IS anomalies. Stimuli under the rubric of Anomaly 1 began with the question *C'est quoi qui se trouve sur la table?* 'What is on the table?'. The cleft construction *c'est quoi qui* places focus on *quoi* 'what', leading to the expectation that the constituent whose referent is co-indexed with *quoi* will also be a focal element in the response. In the control stimuli this is the case, as the response contains a felicitous IS configuration (3a); in the other half of the Anomaly 1 stimuli, the IS configuration is inappropriate in that it anomalously places focus on *la table* (3b).

(3) Anomaly 1

a. (Expected IS configuration):

C'est quoi qui se trouve sur la table?

it is what that is located on the table

'What is on the table?'

C'est un marteau qui se trouve sur la table.

it is a hammer that is located on the table

'A hammer is on the table.'

b. (Infelicitous IS configuration):

C'est quoi qui se trouve sur la table?

it is what that is located on the table

'What is on the table?'

#*C'est sur la table que se trouve un marteau.*

it is on the table that is located a hammer

#'A hammer is on the table.'

Anomaly 2 contained questions of the type *Il est où le marteau?* 'Where is the hammer?', in which *où* 'where' is focal. A possible felicitous response is *Le voilà, le marteau* 'There is the hammer', in which the location of the hammer is focal (4a). In the anomalous condition, a cleft construction in the response assigns focus to *le marteau* 'the hammer', the expected topic (4b).

(4) Anomaly 2

a. (Expected IS configuration):

Il est où le marteau?

it is where the hammer

'Where is the hammer?'

- Le voilà, le marteau.*
 it is there the hammer
 ‘There is the hammer.’
- b. (Infelicitous IS configuration):
Il est où le marteau?
 it is where the hammer
 ‘Where is the hammer?’
 #*C’est le marteau qui est là.*
 it is the hammer that is there
 #‘The hammer is there.’

Under Anomaly 3, a brand-new referent was inappropriately assigned topic status. Following the question *Qu’est-ce qu’il y a sur la photo?* ‘What’s in the photo?’, a typical felicitous response would employ a presentational *il y a* construction to introduce a brand-new referent, as in (5a). In the anomalous response (5b), the item in the photo is treated as a topic, which is infelicitous since it ought to be treated as brand-new information according to Prince’s (1981) scale of Assumed Familiarity.

(5) Anomaly 3

- a. (Expected IS configuration):
Qu’est-ce qu’il y a sur la photo?
 what is it that there is in the photo
 ‘What’s in the photo?’
Il y a un marteau qui est sur la table.
 there is a hammer that is on the table
 ‘There’s a hammer on the table.’
- b. (infelicitous IS configuration):
Qu’est-ce qu’il y a sur la photo?
 what is it that there is in the photo
 ‘What’s in the photo?’
 #*Il est sur la table, le marteau.*
 it is on the table the hammer
 #‘The hammer is on the table.’

Stimuli were presented in 10 blocks of 6 exchanges. Each block was presented in conjunction with an onscreen image providing a context for the exchanges. Each image consisted of a photograph of a household object (e.g., a hammer, a hat, an orange) on top of a wooden table. Subjects rated each exchange as acceptable or unacceptable. “Unacceptable” was defined for the subjects as indicating that “the exchange sounds odd, lacks sense or would be very unlikely.”

5.3. Analysis

Correct responses were defined as those where a subject rated a felicitous control sentence as acceptable, or rated a sentence containing an information structure anomaly as unacceptable. Percent-correct scores on the judgment task were calculated for all participants (see Table 1). When the scores for all L2 subjects were pooled together into one aggregate analysis, there was a negative correlation between AoA and task scores that fell just short of significance ($r = -.460$, $p = .063$). This age effect extended across all AoAs, including post-maturational AoA (see Figure 1); such an overarching age effect is consistent with findings from prior age-related research.

Under a strong version of the Critical Period Hypothesis for second language acquisition (e.g., the one discussed in Johnson & Newport 1989), in the presence of a critical period we would expect to see a high level of performance for the earliest arrivals in the L2 environment, followed by a decline in performance and ending with a flattened low level of performance for the later arrivals (i.e., the abstract “stretched Z” geometric representation of a critical period). A visual inspection of Figure 1 shows no such geometry. Instead, there is a cone-shaped distribution pointing downward from ceiling level and widening with later AoA; this increasing variability with increasing AoA is compatible with what is known about cognitive performance with increasing age (e.g., Schaie 1994). While many subjects exhibited ceiling-level performance, the greater number of lower-performing scores for higher-AoA subjects is compatible with a general trend of age effects after maturation. The presence of an unbounded post-maturational age effect extending well into adulthood is indicative of a “stretched 7” geometry, and is therefore not compatible

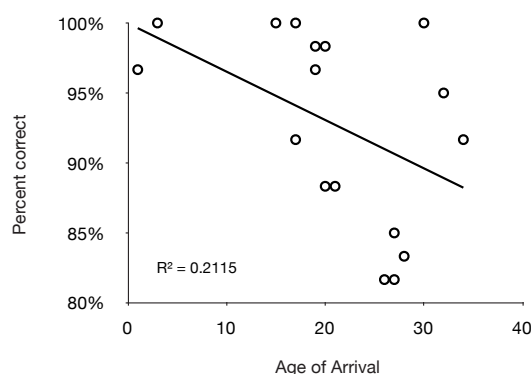


Figure 1. L2 subjects' percent correct scores on judgment task performance versus age of arrival (in years); percent correct reflects scores on all six stimuli types

Table 1. Native speakers' mean scores and L2 subjects' AoA and mean scores on acceptability judgment tasks. Mean score reflects the average scores across all six sentence types tested.

Native speakers Mean score	L2 speakers	
	AoA	Mean score
0.77	1	0.97
1	3	1
0.56	15	1
1		
1	17	0.92
0.99	17	1
1	19	0.97
0.83	19	0.98
0.98	20	0.88
1	20	0.98
1	21	0.88
0.77	26	0.82
0.77	27	0.82
0.98	27	0.85
1	28	0.83
	30	1
1	32	0.95
1	34	0.92
0.81		
0.75		
0.86		
0.93		
1		
0.98		
0.78		
0.76		
1		
0.67		
0.77		
0.98		
0.88		
1		
0.86		
0.92		
0.77		
1		
0.79		
0.86		

Native speakers Mean score	L2 speakers	
	AoA	Mean score
0.89		
1		
0.77		
0.99		
0.95		
0.77		
0.55		

with Johnson & Newport's (1989) stipulation that there must be no ongoing age effects on L2 performance past maturation. Additionally, the presence of ceiling-level performance (including subjects with AoA of as high as 30 scoring perfectly) conflicts with Singleton & Ryan's (2004) requirement that natively-like acquisition is impossible outside the critical period. These results are therefore at odds with Johnson & Newport's (1989) formulation of the Critical Period Hypothesis for the acquisition of these information structural configurations for native speakers of English acquiring French as an L2. However, the correlation between AoA and performance fell short of significance, making any conclusions about age effects somewhat tenuous.

Following Johnson & Newport (1989), DeKeyser (2000), Birdsong & Molis (2001) and others, a disaggregate analysis was conducted in which separate regressions were performed on data from the early and late AoA subjects. By separating the two AoA groups in a disaggregate analysis, we might expect to see clearer evidence of a discontinuity in the AoA function. However, this is not the case. A disaggregate analysis was conducted with the cutoff point between the two AoA groups operationalized as $AoA = 20$.³ The correlations between AoA and task scores did not reach significance for either group: For the early group ($AoA < 20$), $r = -.168$ and $p = .719$. For the late group ($AoA \geq 20$), $r = .095$ and $p = .793$. A visual inspection of the data shows a clear lack of either a "stretched Z" or "stretched L" geometry (see Figure 2). Indeed, the lack of a strong significance for the late arrivals' correlations can be attributed to a high degree of variability in that subset of the data. Note that the data for the late-AoA group does not reflect a flattening of the age function, because the distribution of the data points is random.

3. Disaggregate analyses using earlier cutoff points could not be interpreted in any meaningful way because the number of available data points was too small.

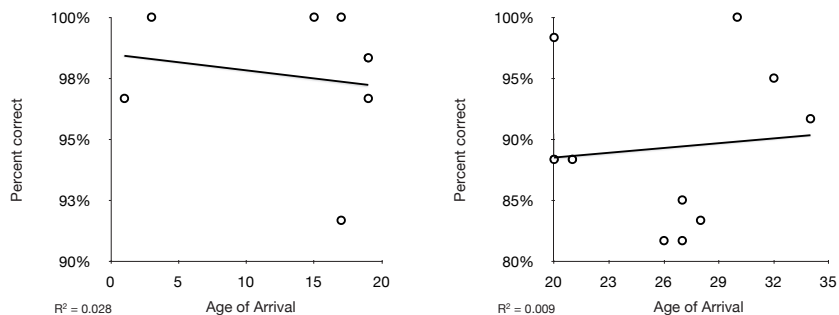


Figure 2. Disaggregate analysis: AoA (years) versus performance on acceptability judgment task. Left: Early group (AoA < 20): $r = -.168, p = .719$. Right: Late group (AoA ≥ 20): $r = .095, p = .793$

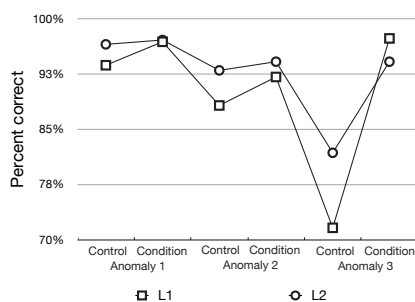


Figure 3. Line plot of percent correct score by sentence type

When control and anomalous conditions of all three IS anomalies were grouped together, the average score was .899 correct for L1 subjects and .922 correct for L2 subjects. A t-test showed a difference in performance between the L1 and L2 subject groups that was not statistically significant ($t = 1.035, df = 59, p = .305$). An analysis of variance (ANOVA) was then performed to look for interaction effects between the subject group and the type of IS anomaly tested. The ANOVA revealed a non-significant interaction of subject group and type of IS construction ($F = 1.873, p = .348, \text{between-subjects } df = 2$). However, a visual inspection of the percent correct scores for each sentence type shows that certain types of stimuli elicited more marked differences between the two subject groups than others (see Figure 3).

The differences between the native speakers' and L2 speakers' scores were measured using t-tests. Scores for each type of information structural anomaly were tested separately. The results of the t-tests showed no significant differ-

ence between the groups' scores for Anomaly 1 ($t = -.183$, $df = 112$, $p = .823$), Anomaly 2 ($t = -.480$, $df = 112$, $p = .189$) or Anomaly 3 ($t = -.414$, $df = 112$, $p = .142$). The interaction line plot in Figure 3 does show some variation in the average scores for the L1 group across the six types of stimuli, most notably for the control sentences of Anomaly 3; however, the difference between the two groups' scores on just the control items fell just short of significance ($t = -.793$, $df = 55$, $p = .07$). I speculate that this result is due to native speakers' intuitions of prescriptive grammar. Several native speaker informants affirmed that the control sentences might be acceptable in spoken French, but not in the written modality, or were otherwise unusual; this concern was addressed in the follow-up experiment (see below).

Correlations between L2 subjects' judgment accuracy (averaged across all stimulus types) and self-ratings of proficiency were calculated (see Table 2). Pearson product-moment correlations were non-significant for self-ratings of vocabulary ($r = .336$, $p = .126$) and syntax ($r = .000$, $p = .998$). The Pearson correlation between accuracy and self-rating of pronunciation fell just short of significance ($r = .470$, $p = .057$); under a Spearman rank-order analysis, this relationship was significant ($Rho = .582$, $p = .014$). These analyses provide mixed evidence for a relationship between an L2 learner's self-assessed proficiency and their actual proficiency in the area of information structure.

To summarize, under the aggregate analysis of the data there was evidence for an AoA effect on task performance across the entire span of AoA. Under a disaggregate analysis of the data, there was no evidence of an age effect geometry consistent with the stretched 7, Z or L representations of the critical period. Instead, there was evidence of ceiling-level performance among early-AoA learners, followed by a random distribution of data points for later AoA learners. Moreover, there were instances of ceiling-level or near-ceiling-level performance among late arrivals, a finding inconsistent with the Johnson & Newport (1989) definition of a critical period.

These findings suggest that L2 acquisition of information structure is not constrained by a critical period. However, the small number of subjects makes it difficult to interpret the disaggregate data. Furthermore, comments from several subjects suggested that normative attitudes towards some of the stimuli influenced their performance on the task, and that the stimuli for Anomalies 2 and 3 were particularly problematic for some subjects. In light of these results, a follow-up experiment was therefore conducted.

6. Experiment 2

In light of the inconclusive results of Experiment 1, a second experiment was conducted to determine if L2 speakers at asymptote perform comparably to na-

Table 2. L2 subjects' self-ratings of proficiency and scores for Experiment 1

AoA	Self-ratings			Schema 1		Schema 2		Schema 3	
	Vocab	Syntax	Pronunciation	Condition	Control	Condition	Control	Condition	Control
1	9	8	9	1	1	1	0.8	1	1
3	7	8	10	1	1	1	1	1	1
15	10	10	10	1	1	1	1	1	1
17	9	9	8	1	1	0.7	1	0.8	1
17	10	10	10	1	1	1	1	1	1
19	8	5	9	1	1	1	1	0.9	1
19	8	6	9	1	1	1	0.9	1	0.9
20	9	9	9	1	1	1	1	1	0.9
20	10	10	9	1	1	1	1	0.9	0.4
21	8	8	8	1	1	0.4	1	0.9	1
26	7	9	9	1	0.9	1	1	1	0
27	7	6	5	0.5	1	0.9	1	0.6	0.9
27	9	9	5	1	1	1	1	1	0.1
28	4	7	2	1	0.7	1	0.7	0.9	0.7
30	7	7	3	1	1	1	1	1	1
32	7	5	9	1	1	1	0.7	1	1
34	8	7	6	1	0.8	1	0.7	1	1

tives on IS judgment tasks; the new experiment was also intended to address the ease of the task in Experiment 1. Experiment 2 consisted of the behavioral component of an experiment involving the measurement of electrophysiological data (Reichle 2008) in which L1 and L2 speakers of French performed an acceptability judgment task for sentences instantiating felicitous and infelicitous information structures. L2 subjects were those whose Age of Exposure (AoE) to the target language was around the end of the putative critical period, with varying levels of L2 proficiency.

In Experiment 1, the scores among L2 subjects were at or near ceiling, and Experiment 2 was intended to address this concern. One possible reason for the high scores in Experiment 1 is that all L2 subjects were of a high enough proficiency that the task presented minimal difficulty; manipulating proficiency as a learner variable allows us to see whether an increase in proficiency leads to an improvement in performance. Conversely, if the task is uniformly easy for all L2 learners and we see no correlation between proficiency and task scores, this would suggest that making information structure judgments is inherently easy for L2 learners with English as an L1, perhaps due to transfer from L1 English which instantiates cleft structures comparable to those in French (though they are used in marked contexts). In Experiment 2, the stimuli were limited to sentences similar to those seen for Anomaly 1 in Experiment 1, taking into account comments from several subjects who found Anomalies 2 and 3 problematic. Finally, controlling the rate at which sentences were presented to participants (see Section 6.2) eliminated the possibility of subjects' deliberating over their judgments for an extended period, and presumably forced a more spontaneous response.

6.1. *Subjects*

Control subjects were adult native speakers of French ($N = 12$). The two experimental groups consisted of low-proficiency L2 speakers of French ($n = 12$) and high-proficiency L2 speakers of French ($n = 12$), all of whom spoke English as an L1. Subjects were recruited from the faculty and student population of the University of Texas at Austin, as well as from the local francophone community. L2 subjects were assigned to a proficiency group based on the composite average of self-reported speaking, reading, listening and pronunciation levels, as measured by a modified version of the LEAP-Q questionnaire (Marian et al. 2007). Subjects were closely matched for Age of Exposure (AoE) between the two L2 groups in order to control AoE as a variable (see Table 3). The average AoE for the low-proficiency subjects was 16.4 years, and the average AoE for the high-proficiency subjects was 15.4 years. The difference in mean AoE between the two groups was non-significant ($F = .016$,

Table 3. L2 subject age of exposure (AoE) in years to the target language

	High-proficiency L2	Low-proficiency L2
	11	12
	12	12
	12	13
	13	14
	14	15
	14	15
	15	18
	16	18
	18	18
	18	18
	19	20
	23	24
Average	15.4	16.4

$df = 22$, $p = .902$). The average time spent in the target environment for the low-proficiency subjects was 3.5 months, while the average time spent in the target environment for the high-proficiency subjects was 1 year. Since the low-proficiency subjects had, on average, spent very short periods of time in the L2 environment, it was determined that the AoA variable was not meaningful for these subjects. Therefore, AoE (defined as the age of first classroom exposure to the L2) was considered for all L2 subjects rather than AoA.

All subjects were at least 18 years of age, and were right-handed. The average age of subjects was 31.4 years for the L1 subjects (range = 19–48), 21.5 years for the low-proficiency L2 subjects (range = 20–30), and 27.6 years for the high-proficiency L2 subjects (range = 18–41).

6.2. Procedure

Subjects were seated in front of a monitor in a quiet booth and were visually presented with a photograph of a household item on a table, followed by a question and response relating to the contents of the photograph. The target sentences (i.e., the response portions of the brief dialogues) were presented one word at a time, with each word appearing at the center of the screen for 300 ms. The interval between the onset of each word in the target sentence was 650 ms.⁴ Using a response box, subjects indicated whether they thought the target

4. These timing intervals were chosen due to constraints from the ERP data collection process

sentences would be acceptable in spoken French. An acceptable sentence was defined for the subjects as “one that they could imagine a person saying that would not seem unusual” given the preceding context.

Before presentation of stimuli, subjects were shown a list of 10 decontextualized sentences representative of spoken French (e.g., *Où que tu vas?* ‘Where are you going?’, *Y a le téléphone qui sonne.* ‘The phone’s ringing’, *T’as mal au genou? Non, c’est mon pied qui me fait mal.* ‘Your knee hurts? No, it’s my foot that hurts’). Subjects were informed that the experiment was meant to reflect spoken French (despite consisting of written sentences), and they were asked to verify that they could imagine situations in which these decontextualized example sentences could be spoken, despite the fact that they are usually not written. This was intended to counteract the influence of normative attitudes toward the written modality, in light of the results of Experiment 1.

Subjects were presented with 55 control sentences (6), 55 condition sentences containing an infelicitous IS configuration (7), and 90 filler sentences.⁵ The control and IS condition sentences were comparable to those used for Anomaly 1 in Experiment 1.

- (6) Control (expected IS configuration – informational focus):

C’est quoi qu’on voit sur la table?
 ‘What do we see on the table?’
C’est un marteau qu’on voit sur la table.
 ‘We see a hammer on the table.’

- (7) Condition (infelicitous IS configuration):

C’est quoi qu’on voit sur la table?
 ‘What do we see on the table?’
 #*C’est sur la table qu’on voit un marteau.*
 #‘We see a hammer on the table.’

Filler sentences were included to prevent the subjects from guessing the true nature of the experiment. Half of the filler sentences contained a target sentence that was identical to that of the corresponding control sentence. However, the preceding context was manipulated by introducing a choice between two referents, thus making the *c’est* cleft in the response an instantiation of contrastive focus (8).

that was taking place simultaneously as part of another study.

5. As mentioned above, data collection was conducted in conjunction with an electrophysiological recording session. In several cases, data collection ended prior to the presentation of all stimuli sentences due to constraints from the electrophysiological recordings.

- (8) Filler 1 (expected IS configuration – contrastive focus):
Est-ce que c'est un verre ou un marteau qu'on voit sur la table?
 'Is it a glass or a hammer that we see on the table?'
C'est un marteau qu'on voit sur la table.
 'We see a *hammer* on the table.'

The other filler sentences contained a semantic mismatch. Target sentences named an object whose identity conflicted with that of the object in the image the subject had just seen (9):

- (9) Filler 2 (expected IS configuration – semantic mismatch):
C'est quoi qu'on voit sur la table?
 'What do we see on the table?'
 **C'est un journal qu'on voit sur la table.*
 *'We see a *newspaper* on the table.'

6.3. Analysis

Acceptability judgments for the infelicitous IS condition sentences were tabulated for the three subject groups (see Table 4). For the control sentences and both types of filler sentences, subjects in all three groups judged nearly all stimuli accurately. Here 'accurately' means that they judged felicitous control sentences as acceptable, and rejected anomalous sentences as unacceptable. Judgments of the IS anomalies, on the other hand, departed from ceiling, with L1 subjects accurately judging 47.3% of these anomalies, low-proficiency L2 subjects judging 24.9% accurately, and high-proficiency L2 subjects judging 53% accurately.

A look at the responses for the IS anomaly sentences shows that subjects in all three groups displayed a preference to consistently classify the anomalous IS sentences as either all being acceptable, or all being unacceptable (see Table 5). Overall, the tendency was for subjects to judge the anomalous sentences as acceptable, i.e., most of the subjects made inaccurate judgments.

Table 4. *Accurate judgment scores by condition*

	Control %	IS anomaly %	Felicitous filler %	Semantic mismatch filler %
L1	98.9	47.3	97.6	95.8
LL2	99.5	24.9	98.7	96.7
HL2	98.4	53.0	96.5	98.3

Table 5. Accurate and inaccurate responses for acceptability judgment of IS anomaly sentences (Responses belonging to subjects who accurately judged the IS condition sentences to be unacceptable are in bold)

AoE	Low-proficiency L2		AoE	High-proficiency L2		Native speakers	
	Accurate	Inaccurate		Accurate	Inaccurate	Accurate	Inaccurate
12	48	1	11	50	0	54	1
12	41	9	12	1	49	49	1
13	3	46	12	20	21	48	2
14	0	55	13	1	40	47	3
15	1	49	14	45	7	41	9
15	1	49	14	0	32	29	1
18	0	32	15	55	0	1	39
18	54	1	16	47	2	0	49
18	0	50	18	50	2	0	43
18	2	28	18	1	54	0	50
20	2	48	19	1	49	0	43
24	0	37	23	46	0	0	50
	152	405		317	256	269	291

A minority of subjects in each group accurately rejected the anomalous sentences as unacceptable. With the exception of one high-proficiency L2 subject (AoE = 12) who was inconsistent in his judgments, rarely did an individual judge the sentences in a way that deviated from his overall preference. When percent-correct scores were calculated for each subject, this revealed bimodal distributions within each of the three subject groups: Each group contained some subjects who overwhelmingly judged the IS condition sentences to be acceptable, and several other subjects who judged them to be unacceptable.

The responses shown in bold are those of subjects who consistently judged the IS anomaly sentences to be unacceptable. As seen in Table 5, only three low-proficiency L2 subjects displayed such behavior, while six high-proficiency L2 subjects and six L1 subjects displayed the same behavior. The fact that only three low-proficiency subjects made accurate judgments, compared to six subjects in each of the other proficiency groups, suggests some difference in task performance due to differences in proficiency. A chi-square analysis of the raw scores confirms that the different proficiency groups performed differently: The scores for the low-proficiency group diverged significantly from those of the L1 group (chi-square = 51.2, $df = 1$, $p < .001$) and the high-proficiency L2 group (chi-square = 91.4, $df = 1$, $p < .001$). The scores for the L1 and high-proficiency L2 groups also diverged significantly (chi-square = 6.02, $df = 1$, $p = .014$); however, the fact that six out of twelve subjects in both groups judged the IS anomalies to be unacceptable indicates that these two groups behaved similarly in this respect. This similarity between the performance of the L1 and high-proficiency L2 subjects is interpreted as modest evidence for an increase in judgment task ability with an increase in L2 proficiency.

At the same time, the L1 and high-proficiency L2 groups appear to have scored near chance levels. A chi-square analysis comparing the observed responses with those expected under chance conditions shows that the L1 subjects (chi-square = .280, $df = 1$, $p = .597$) and high-proficiency L2 subjects (chi-square = 3.25, $df = 1$, $p = .071$) scored at chance levels. The low-proficiency subjects, most of whom judged the anomalous sentences incorrectly, did not score at chance levels (chi-square = 60.5, $df = 1$, $p < .001$). Importantly, we cannot attribute their scores to a general response bias to accept all stimuli: The same subjects accurately identified the control and felicitous filler sentences as acceptable, and accurately identified the semantic mismatch filler sentences as unacceptable (See Table 4). A chi-square analysis comparing the observed responses with those expected under an incorrect response bias confirms this (chi-square = 176, $df = 1$, $p < .001$). Given these results, any given L1 or high-proficiency L2 subject has an approximately 50–50 chance of correctly judging the anomalies, whereas the majority of low-proficiency L2 subjects will judge the same anomalies inaccurately.

To shed light on the reasons for these bimodal distributions, behavioral data was integrated with electrophysiological data recorded during the same task as part of another experiment (Reichle 2008). Each subject group was divided into two subgroups: one consisting of subjects who responded accurately to the IS anomaly judgment task, and another consisting of subjects who judged the same anomalies inaccurately.

Univariate ANOVAs were conducted to look for significant differences between the correct and incorrect responder subject groups. The dependent variable was the mean amplitude of the electrophysiological waveforms within a given period of time post-stimulus; these values were compared between the control and IS conditions (see Table 6). One major caveat is that the utility of such ANOVAs is limited due to the small size (and therefore, lack of normal distribution) of the subgroups – the largest subgroup had only six members. However, it is useful as a means of examining the differences between the responders.

Previous ERP research on native speaker subjects has established that a negative increase in voltage around 400 ms post-stimulus is indexed with the processing of semantic anomalies, e.g., errors in truth value (Kutas & Hillyard 1980). This pattern of electrophysiological activity is referred to as the N400 effect. For the processing of syntactic errors, on the other hand, a positive increase in voltage around 600 ms post-stimulus, referred to as the P600 effect, has been found (Osterhout and Holcomb 1992). Given that the previously established N400 and P600 effects occur in the 300–500 ms and 500–700 ms time windows, respectively, ANOVAs were conducted on the mean amplitude data of the six subgroups in these time windows.

The analysis of the L1 subgroups shows a significant interaction between anteriority and subgroup in the 300–500 ms time window ($F = 19.248$, $df = 1$, $p = .036$), and no significant effects in the 500–700 ms time window. An inspection of the topographic distribution of the ERPs showed a right anterior and posterior negativity from 400–500 ms for the correct responders; for incorrect responders, there was a left posterior negativity in the same window. For the low-proficiency L2 subgroups, there was a significant interaction between anteriority and subgroup in the 300–500 ms window ($F = 11.818$, $df = 1$, $p = .033$), and a marginal effect for subgroup ($F = 32.57$, $df = 1$, $p = .064$) and a significant interaction between anteriority and subgroup ($F = 12.146$, $df = 1$, $p = .026$) in the 500–700 ms window. This was seen as an increased anterior positivity for the correct responders in both the 300–500 ms and 500–700 ms time windows. For the high-proficiency L2 subgroups, a significant interaction between hemisphere and subgroup ($F = 26.082$, $df = 2$, $p = .037$) was observed in the 300–500 ms window; no significant effects were seen in the 500–700 ms window. This was seen as an increase in positivity in the left hemisphere for incorrect responders.

Table 6. Summary of ANOVA tests of effects between correct and incorrect responder subgroups. Significant effects are in bold. Subgroup = correct vs. incorrect responders; HEM = hemisphere; ANTER = anteriority; LL2 = low-proficiency L2; HL2 = high-proficiency L2.

Factor	Subject group	Time window (ms)	<i>p</i>	<i>df</i>	<i>f</i>
Subgroup	L1	300–500	0.733	1	0.193
Subgroup * HEM	L1	300–500	0.357	2	1.804
Subgroup * ANT	L1	300–500	0.036	1	19.248
Subgroup	L1	500–700	0.669	1	22.999
Subgroup * HEM	L1	500–700	0.995	2	0.005
Subgroup * ANT	L1	500–700	0.331	1	1.549
Subgroup	LL2	300–500	0.164	1	9.023
Subgroup * HEM	LL2	300–500	0.254	2	2.937
Subgroup * ANT	LL2	300–500	0.033	1	11.818
Subgroup	LL2	500–700	0.064	1	32.57
Subgroup * HEM	LL2	500–700	0.189	2	4.28
Subgroup * ANT	LL2	500–700	0.026	1	12.146
Subgroup	HL2	300–500	0.352	1	1.517
Subgroup * HEM	HL2	300–500	0.037	2	26.082
Subgroup * ANT	HL2	300–500	0.921	1	0.01
Subgroup	HL2	500–700	0.509	1	0.653
Subgroup * HEM	HL2	500–700	0.105	2	8.564
Subgroup * ANT	HL2	500–700	0.464	1	0.715

The fact that correct and incorrect responders exhibited different electrophysiological correlates for the processing of IS anomalies suggests the possibility of differences in processing. It may be, for example, that the L1 subjects who accurately judged the anomaly sentences to be unacceptable processed the sentences in ways similar to how they would process truth-value errors (hence the similarity of their processing signatures to the N400 component), whereas the subgroup that inaccurately judged the anomalies showed processing signatures closer to the LAN component associated with syntactic anomalies. Similarly, the high-proficiency L2 subjects who accurately judged the stimuli did not exhibit the same increase in positivity seen among subjects who judged the same anomalies inaccurately; this suggests that the correct responders exhibited processing signatures that approximated those seen among correct-responding L1 subjects, while the signatures of the incorrect responders were dissimilar in location and polarity from those of L1 subjects. Notably, this implies that different segments of these populations use different processing techniques when presented with IS anomalies. (This is

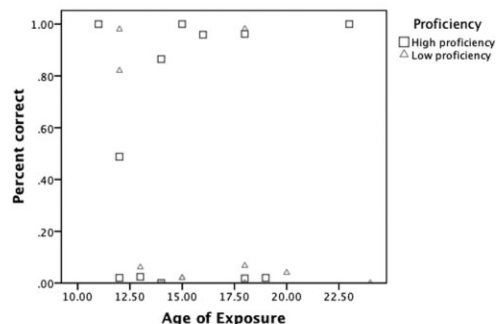


Figure 4. Percent-correct scores versus age of exposure to the target language (in years) on anomalous IS condition judgment task

independent from the question of how they process IS under *felicitous* conditions.) Unfortunately, the small sizes of these subgroups make it impossible to draw any strong conclusions from these results, and further studies would need to be done in which subjects' behavioral tendencies were taken into account when recruiting subjects for the recording of electrophysiological data.

An examination of percent-correct scores on the anomalous IS condition judgment task and initial age of exposure to French shows no relationship between the two variables (Figure 4). A visual inspection of the data shows that earlier learners of French did not necessarily perform better on the task than later learners. This applies to both high- and low-proficiency subjects. The Pearson correlation between AoE and task score was non-significant for low-proficiency subjects ($r = -.469$, $p = .124$), high-proficiency subjects ($r = .157$, $p = .625$), and for the two L2 groups combined ($r = -.141$, $p = .511$).

A disaggregate analysis of the L2 subjects' scores was also conducted, in which subjects of $AoE \leq 14$ were grouped together as early exposure learners, and subjects of $AoE > 14$ were grouped together as late exposure learners. There was no significant correlation between AoE and task scores for either the early exposure learners ($r = -.506$, $p = .135$) or the late exposure learners ($r = -.061$, $p = .836$).

No significant correlation was found between judgment task scores and subjects' averaged speaking, reading, listening and pronunciation self-ratings ($r = .077$, $p = .720$), nor was there any relationship between judgment task scores and the amount of time spent in a Francophone country ($r = -.234$, $p = .464$ for the low-proficiency group; $r = .285$, $p = .370$ for the high-proficiency group; $r = .224$, $p = .293$ for both L2 groups combined).

The results of the chi-square analysis (above) indicate that, as a group, high-proficiency L2 speakers performed better at the information structure judgment task than low-proficiency L2 speakers did. However, an examination of Table 5 suggests an extremely high degree of variability from subject to subject in terms of predictiveness of identified factors.

Furthermore, the lack of consensus regarding the acceptability of the anomalous IS sentences, even among the L1 subjects, gives pause. What truly makes an utterance acceptable or unacceptable to these subjects? There are two possible sources for this inter-subject variability: One is that the operationalization of acceptability (a sentence “they could imagine a person saying, and it would not seem unusual”) that was provided to the subjects lent itself to more than one interpretation. To avoid this in future studies, the definition of acceptability should be very precise. The other possible source of variability in subjects’ responses is that, for some speakers, the target sentence’s referentiality and truth value overshadowed the influence of the information structure of the preceding context during the judgment task. I consider this a likely source of variability, especially in light of comments several subjects made after the session. These subjects stated that their knowledge of the truth value of the target sentence conflicted with what they knew about the context in which it appeared, and that this made them biased towards responding based on truth value (e.g., when presented with *#C’est sur la table qu’on voit un marteau*, the subject rated the sentence as acceptable despite its anomalous focus structure because the sentence was true inasmuch as the photo did contain a hammer on the table).

7. Conclusion

The results of these experiments provide evidence regarding the existence of a putative critical period for the L2 acquisition of information structure, and about the way information structure is processed.

Two findings act as counter-evidence to the claims of the strong version of the Critical Period Hypothesis for second language acquisition:

- (1) Geometric evidence: Under an aggregate analysis of the data from Experiment 1, there is marginal evidence for a decline in L2 speakers’ ability to judge the acceptability of ill-formed information structural configurations as age of arrival in the L2 environment increases. The relationship between judgment task proficiency and age of arrival continues to decline past the years of physical and cognitive maturation and well into adulthood, unlike the “stretched Z” geometry of a putative critical period (a criterion of Johnson & Newport 1989). Since the geometry of this correlation indicates that post-maturational age effects are observed for L2 speakers, this is at

odds with the predictions of a Johnson and Newport's (1989) version of the Critical Period Hypothesis for second language acquisition regarding the acquisition of information structure in French.

- (2) Native-like performance: In both experiments, the high level of native-like performance on the judgment task among late arrivals in the L2 environment is contrary to one of the criteria for a critical period. L2 subjects in both experiments showed native-like performance on information structure judgment tasks. This violates Johnson & Newport's (1989) criterion that late-AoA learners should not exhibit native-like behavior, as well as Singleton & Ryan's (2004) criterion that learners cannot acquire a target behavior once a critical period has passed. In this experiment, the high degree of nativelikeness may not be suitable evidence for rejection as the judgment task may have been too easy for many of the respondents (see Long 1990). Relatedly, we must take into account the possibility of transfer between English as an L1 and French as an L2 as far as information structure is concerned. This possibility warrants further investigation. However, the fact that the group of classroom learners in Experiment 2 who had little to no experience in the target environment did not exhibit nativelike performance suggests that transfer is not taking place, since any such transfer effect should apply equally to all L2 learners regardless of their time spent in the target environment.

Regarding the way in which information structure is processed, the results of Experiment 2 suggest that under controlled experimental conditions, for some subjects truth value takes precedence over focus structure when judging acceptability, but for other subjects the reverse is true. This was observed in the behavioral data and electrophysiological data for both native and advanced L2 speakers: subjects who judged stimuli based on truth value exhibited different electrophysiological signatures from subjects who judged the same stimuli based on focus structure. Because of the small amount of available data for these behavioral subgroups, further studies with more participants should be conducted to examine these differences in processing.

While there was a significant positive correlation between self-rating of pronunciation and task performance in Experiment 1, there were no correlations between proficiency self-ratings and performance in Experiment 2, nor was there a relationship between age of exposure and task score. Correlations were found between performance and AoA and between performance and proficiency (Experiment 1), but not between performance and AoE or time spent in the target country (Experiment 2). While differences in the factors considered by the two experiments make direct comparison of results impossible, it may be that a high level of performance in judging information structure is possible due to early acquisition or a high level of proficiency, but not from classroom exposure or shorter stays in the target environment. (Recall that the average

length of residence in Experiment 1, where a correlation was found, was 10 years, versus the 1-year average length of residence for high-proficiency speakers in Experiment 2.) Additional studies examining L2 learners with a wider range of lengths of residence would clarify the nature of this relationship.

Northern Illinois University
<rreichle@niu.edu>

References

- Bialystok, Ellen. 2002. On the reliability of robustness: A reply to DeKeyser. *Studies in Second Language Acquisition* 24(3). 481–488.
- Bialystok, Ellen & Kenji Hakuta. 1999. Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In David Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, 161–181. Mahwah, NJ: Erlbaum.
- Birdsong, David. 2005. Interpreting age effects in second language acquisition. In Judith F. Kroll and Annette M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 109–127. New York: Oxford University Press.
- Birdsong, David. 2006. Age and second language acquisition and processing: A selective overview. *Language Learning* 56(s1). 9–49.
- Birdsong, David & Michelle Molis. 2001. On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44. 235–249.
- Bley-Vroman, Robert. 1989. What is the logical problem of foreign language learning? In Susan M. Gass and Jacquelyn Schacter (eds.), *Linguistic perspectives on second language acquisition*, 41–68. Cambridge: Cambridge University Press.
- Bornstein, Marc H. 1989. Sensitive periods in development: Structural characteristics and causal interpretations. *Psychological Bulletin* 105. 179–197.
- Carroll, Mary & Monique Lambert. 2003. Information structure in narratives and the role of grammaticised knowledge: A study of adult French and German learners of English. In Christine Dimroth and Marianne Starren (eds.), *Information structure and the dynamics of language acquisition*, 267–288. Amsterdam: John Benjamins.
- Carroll, Mary, Jorge Murcia-Serra, Marzena Watorek, & Alessandra Bendiscioli. 2000. The relevance of information organization to second language acquisition studies: the descriptive discourse of advanced adult learners of German. *Studies in Second Language Acquisition* 22. 441–466.
- Carroll, Mary & Christiane von Stutterheim. 2003. Typology and information organisation: Perspective taking and language-specific effects in the construal of events. In Anna Giacalone Ramat (ed.), *Typology and Second Language Acquisition*, 365–402. New York: Walter de Gruyter.
- Colombo, John. 1982. The critical period concept: Research, methodology, and theoretical issues. *Psychological Bulletin* 91. 260–275.
- DeKeyser, Robert. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22(4). 499–533.
- Dimroth, Christine. 2008. Age effects on the process of L2 acquisition? Evidence from the acquisition of negation and finiteness in L2 German. *Language Learning* 58(1). 117–150.
- Dimroth, Christine, Petra Gretsche, Peter Jordens, Clive Perdue, & Marianne Starren. 2003. How to express finiteness in Germanic languages: A stage-model for first and second language acquisition. In Christine Dimroth and Marianne Starren (eds.), *Information structure and the dynamics of language acquisition*, 65–94. Amsterdam: Benjamins.

- Ferdinand, Astrid. 2002. Acquisition of syntactic topic marking in L2 French. *Linguistics in the Netherlands* 19(1). 49–59.
- Fllege, James, Grace Yeni-Komshian, and Serena Liu. 1999. Age constraints on second language acquisition. *Journal of Memory and Language* 41. 78–104.
- Hakuta, Kenji, Ellen Bialystok, & Edward Wiley. 2003. Critical evidence: a test of the critical-period hypothesis for second-language acquisition. *Psychological Science* 14(1). 31–38.
- Johnson, Jacqueline S. 1992. Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning* 42. 217–248.
- Johnson, Jacqueline & Elissa Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a Second Language. *Cognitive Psychology* 21. 60–99.
- Johnson, Jacqueline S. & Elissa L. Newport. 1991. “Critical period effects on universal properties of language: The status of subadjacency in the acquisition of a second language.” *Cognition* 39. 215–258.
- Kaiser, Elsi & John C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94(2). 113–147.
- Kraut, Robert, Judith Olson, Mahzarin Banaji, Amy Bruckman, Jeffrey Cohen, & Mick Couper. 2004. Psychological research online: Report of Board of Scientific Affairs’ Advisory Group on the Conduct of Research on the Internet. *American Psychologist* 59(2). 105–117.
- Kutas, Marta & Steven A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science* 207. 203–205.
- Lambrecht, Knud. 1994. *Information structure and sentence form: A theory of topic, focus, and the mental representations of the discourse referents*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3). 463–516.
- Lambrecht, Knud, Julia Bordeaux, and Robert Reichle. 2006. Cognitive constraints on assertion scope: The case of spoken French *Parce que*. In Chiyo Nishida and Jean-Pierre Y. Montreuil (eds.), *New Perspectives on Romance Linguistics: Vol. I: Morphology, Syntax, Semantics, and Pragmatics. Selected papers from the 35th Linguistic Symposium on Romance Languages (LSRL), Austin, Texas, February 2005*, 143–154. Amsterdam: John Benjamins.
- Lenneberg, Eric. 1967. *Biological foundations of language*. New York: Wiley.
- Long, Michael H. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12. 251–285.
- Marian, Viorica, Henrike K. Blumenfeld, and Margarita Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research* 50. 940–967.
- Newport, Elissa L. 1991. Contrasting concepts of the critical period for language. In Susan Carey and Rochel Gelman (eds.), *The Epigenesis of mind: essays on biology and cognition*, 111–132. Hillsdale, N.J.: L. Erlbaum Associates.
- Osterhout, Lee & Phillip J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31. 785–806.
- Pinker, Steven. 1994. *The language instinct: How the mind creates language*. New York: Morrow.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Reichle, Robert V. 2008. *Syntactic focus structure processing: Behavioral and electrophysiological evidence from L1 and L2 French*. PhD Dissertation. University of Texas at Austin. Ann Arbor: ProQuest/UMI.
- Reichle, Robert V. 2009. The critical period hypothesis: Evidence from information structural processing in French. In J. Arabski & A. Wojtaszek (eds.), *Neurolinguistic and psycholinguistic perspectives on SLA*. Bristol, UK: Multilingual Matters.

- Schaie, K. Warner. 1994. Developmental designs revisited. In Stanley Hymie Cohen and Hayne Waring Reese (eds.), *Life-span developmental psychology: Methodological considerations*, 45–64. Hillsdale, NJ: Erlbaum.
- Schlyter, Suzanne. 2003. Development of verb morphology and finiteness in children and adults acquiring French. In Christine Dimroth and Marianne Starren (eds.), *Information structure and the dynamics of language acquisition*, 15–44. Amsterdam: Benjamins.
- Singleton, David. 2005. The Critical Period Hypothesis: A coat of many colors. *IRAL: International Review of Applied Linguistics in Language Teaching* 43(4). 269–285.
- Singleton, David & Lisa Ryan. 2004. *Language acquisition: The age factor*, 2nd Edition. Clevedon: Multilingual Matters.
- Wilson, Frances, Antonella Sorace, & Frank Keller. 2007. L2 processing of information structure: The role of word order and pronominalization in German. Paper presented at International Symposium on Bilingualism 6, University of Hamburg, 30 May – 2 June.