Universal Grammar and Second Language Acquisition: The effect of modality of presentation on a grammaticality judgment task

Victoria A. Murphy

Department of Educational and Counselling Psychology

McGill University, Montréal

August 1993

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master's of Arts in Educational Psychology.

Shortened Title:

The effect of presentation modality on grammaticality judgment tasks.

Abstract

Typical experiments investigating the accessibility and/or role of principles of Universal Grammar (UG) in adult second language acquisition (SLA) use a written grammaticality judgment (GJ) task to infer knowledge of principles of UG. The present investigation examined whether subjects would judge sentences differently in the aural modality than the visual. It was hypothesized that subjects in the aural condition would be less accurate and slower at judging sentences violating the subjacency principle than subjects in the visual condition. Four language groups were tested, ESL (English second language) FSL (French second-language), L1E (English first language) and L1 F (French first language). Subjects were assigned to either an aural or a visual condition; the same sentences were presented via computer. The target sentences presented to the subjects were declarative sentences involving embedded questions, as well as ungrammatical wh questions which violated subjacency. The presentation times for all sentences were matched across conditions. Accuracy and reaction time to grammaticality judgment were measured. The hypothesis that subjects would be slower and less accurate in the aural condition than the visual one was supported. Furthermore, subjects were less accurate and slower to judge violations of subjacency than other sentences, in both modalities. The detrimental effects of the auditory task on judgments was most pronounced for the L2 learners. These results are discussed in the context of the informativeness and validity of outcomes derived from GJ tasks, and the ways in which they are presented.

Résumé

Les expérimentations typiques examinant l'accessibilité et/ou le rôle des principes de *Universal Grammar* (UG) dans l'acquisition d'une deuxième langue à l'âge adult utilisent une tâche de jugement de grammaticalité écrite (JG) afin d'inférer une connaissance des principes de UG La présente investigation a examiné si les sujets jugeraient différement les phrases dans la modalité orale que visuelle. L'hypothèse est que les sujets dans la condition orale seraient moins précis et plus lents à juger les phrases violant le principe de *subjacency* que les sujets dans la condition visuelle. Quatre groupes de languages ont été mis à l'essai; ESL (English second-language), FSL (French second-language), L1.E (English first language) and L1.F (French first language). Les sujets étaient assignés soit à une condition orale ou bien à une condition visuelle; les mêmes phrases étaient présentées par ordinateur. Les phrases cibles présentées aux subjets étaient des phrases déclaratives impliquant des questions enchassées, aussi bien que des questions non grammaticales avec des mots 'wh' qui violent la *subjacency* . Le temps de la presentation des phrases a été agencé aux conditions. La précision et le temps de réaction au jugement de grammaticalité ont été mesurés. L'hypothèse que les sujets seraient plus lents et moins précis dans la condition orale que dans la condition visuelle a été supportée. De plus, les sujets étaient moins précis et plus lents à juger les violations de *subjacency* que d'autres phrases, dans les deux modalités. Les effets nuisibles de la tâche auditoire sur les jugements étaient plus prononcés pour les débutants L2. Ces résultats sont examinés dans le contexte de l'instructivité et la vadilité des conséquences dérivées des tâches de (JG), et de la façon dont ils sont présentés.

Acknowledgements

There are so many people to whom I owe thanks for the completion of this Master's Thesis that I fear I will not be able to include everyone by name. I would first like to thank all of the subjects who participated in my experiment, both the pilot subjects and the ones who participated in the actual experiment. Obviously, without them this thesis would not have been possible. I would also like to thank the members of the Laboratory for Applied Cognitive Science in the Department of Educational Psychology at McGill. Thanks to Hélène Fournier for the translation and Alain Breuleux for providing the computer to run my subjects. My fellow graduate students in this lab were of invaluable assistance to me either by offering me advice along the way, or by answering telephone calls and giving me messages from potential subjects. I thank Dorothy Bartram and the TOEFL office for their assistance in providing me with the TOEFL tests I used on the ESL subjects. I would like to thank José Bonneau from the Department of Linguistics for translating the English stimuli into French for me. Thank you also to Adrien Pitrat who supplied the earphones used for the subjects in the aural condition. I owe special thanks to Frédéric Vallée-Tourangeau for his wonderful native French-speaking voice and his time in making the French stimuli recordings for me. I was fortunate enough to have two supervisors, Dr. Michael Hoover of the Department of Educational and Counselling Psychology and Dr. Lydia White of the Department of Linguistics, guiding me through this project. I am forever grateful for the supervision I received because I know this thesis would not have been successful without it.

I owe the greatest thanks to Robin Murphy. I thank him for his voice which he generously offered for the recording of the English stimuli, but more importantly, for his support in all aspects of my life. It is due to his unwavering encouragement and support of me that I am capable of accomplishing anything, and I consider myself undeservingly fortunate to have him in my life.

Table of Contents

Universal Grammar and Second Language Acquisition.
The effect of modality of presentation on a grammaticality judgment task

1.0 Introduction

The accessibility and role of Universal Grammar (UG) in second language acquisition (SLA) has yet to be determined. Since Chomsky's (1981a) Government and Binding (GB) theory, researchers have enthusiastically attempted to apply GB theory to the investigation of adult SLA. GB theory assumes that principles and parameters of UG constitute an innate body of knowledge which constrains first language (L1) acquisition (White, 1989). The principles are universal linguistic properties found across all languages, while the parameters identify the extent of variation across languages. With the specification of some of the constraints which might be operating in language acquisition, researchers were able to investigate whether these universal principles and parameters are accessible to adult second language (L2) learners (White, 1988, Bley Vroman, Felix & Ioup, 1988; Schachter, 1989). However, despite over a decade of research using different designs, investigating different languages and using different subject populations, reliable and convincing evidence concerning UG's accessibility and/or role in SLA has yet to be accumulated.

A probable account of why one as yet cannot convincingly argue for or against a role for UG in adult SLA may centre around methodological issues. Researchers, such as Chaudron (1988), Bley-Vroman & Masterson, (1989) and Birdsong, (1989, 1992), have explicitly outlined how methodological issues such as response biases, validity of measurement and problems with data collection can and do lead to inconclusive research if not dealt with properly. Nevertheless, these methodological concerns have often been neglected, as is evident in much of the research conducted to date (see for example, Schachter, 1989, Clahsen & Muysken, 1986). Researchers must take methodological issues into consideration since their neglect can lead to research that is questionable and/or conflicting (for example, Schachter, 1989 with White, 1988).

Schachter (1989) and White (1988) each reported an investigation into L2 learners' knowledge of the principle of subjacency. Schachter's (1989) subjects were learners English whose L1 was either Chinese, Korean, and Indonesian. White's (1988) subjects were also learners of English whose L1 was French. Though both White's and Schachter's investigations involved L2 learners of English, used a grammaticality judgment (GJ) task, and investigated the same universal principle (subjacency), they reached different conclusions. Namely, White (1988) argues that subjacency is accessible to L2 learners, regardless of the L1, while Schachter (1989) suggests that subjacency is only accessible in the L2 via transfer from the L1.

Both the White (1988) and Schachter (1989) experiments used a GJ task in investigating knowledge of universal principles. The GJ task is a prevalent means of measuring linguistic competence in SLA. GJ tasks are ones in which sentences are presented to subjects who are then asked to judge whether or not the sentences are correct, grammatical, good and so forth. There are both advantages and disadvantages to using a GJ task. One obvious advantage is that such a task allows experimenters to construct the sentences in such a way that the specific principle (or linguistic phenomenon) under investigation will be included. If the experimenter relies on spontaneous production data for example, the subjects may never produce the structure the experimenter is interested in. Therefore, GJ tasks allow for a degree of control over the experiment since subjects are forced to consider the specific structure under investigation, and the experimenter can include violations of that structure (White, 1989). There are problems, however, in using a GJ task.

Birdsong (1989) argues that GJ tasks can be somewhat limited as to their informativeness because subjects could be exhibiting a response bias in judging the sentences. Birdsong (1989) also suggests that just because a subject correctly judged a sentence does not guarantee that the subject recognizes the structure being manipulated. For example, if a subject were to judge correctly a violation of a universal principle as

ungrammatical, the experimenter cannot be sure that this subject has knowledge about that particular universal principle. Subjects may not attend to syntactic variables when doing a GJ task but may indeed be attending to semantic criteria (Birdsong, 1989, Kellerman, 1985). Furthermore, a subject may simply not be capable of the metalinguistic skills necessary to perform successfully on a GJ task.

The problems of using GJ tasks, however, does not mean that one should abandon such a measure. Some of the problems can be avoided. If subjects are required to judge both grammatical and ungrammatical sentences, then successful performance is unlikely to be due to a response bias. Furthermore, using linguistically naive subjects (i.e., not linguistic students or professors) in order to obtain 'true intuitions' (Birdsong, 1989) and acquiring proficiency measures to obtain homogenous groups of subjects can be useful in enabling experimenters to conduct sound research.

A methodological issue central to the present study, which has only recently attracted attention, is the modality of stimulus presentation (written versus aural) in GJ tasks. Modality is a variable which should be considered since a difference in modality can lead to differential responding where subjects may be more accurate, or respond more quickly, in one modality than in another (Johnson, 1992, Haig, 1991).

This thesis will conform to the following structure, the implications of UG in general, and specifically to SLA will be considered. A discussion of specific methodological problems will follow to demonstrate the validity of the claim that methodology can and does detrimentally affect the outcomes of research. The issue of modality of stimuli presentation will then be discussed to demonstrate how modality is a methodological concern that can be controlled and improved upon. The experimental investigation undertaken in this thesis is then described and discussed as an example of research clearly demonstrating the effects of modality on subjects' responding on a grammaticality judgment (GJ) task.

1.1 Universal Grammar and Second Language Acquisition

1.1.1 Universal Grammar

Chomsky stated "Universal grammar may be thought of as some system of principles, common to the species and available to each individual prior to experience" (Chomsky, 1981b, p 7). Chomsky's (1965, 1981a) proposal of UG suggested that children have an innate knowledge of certain universal linguistic principles. UG then is a species specific capacity which is the basis for acquiring language. Chomsky's notions of language acquisition were particularly provocative because they overthrew the previously held belief that language was an arbitrary system, learned by principles of repetition and reinforcement (Skinner, 1957).

The development of the theory of UG was motivated by the observation that the linguistic input children receive underdetermines their final knowledge of language. Children hear a finite set of sentences from their parents or guardians while they are learning their language, yet natural language is an open-ended set of sentences (Pinker, 1989). Additionally, children do not get information about structures that are ungrammatical in their language (negative evidence) but rather, only hear structures which are licensed by their language (positive evidence). This lack of negative evidence is a significant problem to overcome in explaining the child's successful acquisition. Without negative evidence a child could conceivably acquire a grammar that allows too many structures, in essence it would over generate (Pinker, 1990). For example, if children do not receive negative evidence, they might never learn that forms such as 'Who does John like Mary and?' are not grammatical in English. Most research investigating the issue of negative evidence to date suggests that children do not receive negative evidence in any substantial form (see Pinker, 1990 for a brief review). The logical problem of language acquisition is that the child's environment does not tell him/her which forms are grammatical and which are not, yet the child is easily able to attain a grammar which clearly integrates information about grammaticality.

The theory of UG attempts to describe how children learn which linguistic structures are grammatical and which are not by accounting for the gap between the linguistic experience a child receives and the attained linguistic competence the child will acquire. This gap is accounted for by providing the child with knowledge of language before the child even begins to learn, i.e., 'built-in' knowledge. This knowledge is characterized by principles of UG which determine the form of any grammar Immediately, therefore, the child's task is constrained because the grammar will adhere to constraints outlined by UG Language specific knowledge is acquired through the aid of these principles which the child already has and does not need to learn Therefore, certain errors will not occur in child speech since the child already tacitly knows something about language (i.e., what the form of his/her grammar could possibly look like).

The innate knowledge of universal principles by itself is not enough to be able to learn a language. The linguistic environment within which a child develops will determine which language is learned. A child growing up in a French speaking environment for example will not become a native speaker of Russian While some aspects of human languages are universal, there is also much variation. There are language-specific components to language learning, like word order (Haegeman, 1991) In UG, as identified by GB theory, this language-specific variation is characterized by parameters. The child has to determine what the setting of a particular parameter is For example, the English child has to learn that the word-order parameter in English is set for head initial, yielding subject-verb-object. The Japanese child, however, has to learn that the word-order parameter in Japanese should be set to head final, giving subject-object-verb (Haegeman, 1991). Similarly, other aspects of languages are subject to parametric variation.

When learning their language, children use their innate knowledge of universal principles as characterized by UG in addition to the input received in their linguistic

environment, which informs them of the setting for a specific parameter (Haegeman, 1991).

Subjacency is an example of a principle of UG which constrains syntactic movement in language. Linguistic constituents are regularly 'moved' from the deep structure position to yield a grammatical surface structure representation. Subjacency is a condition which restricts movement of linguistic entities in certain ways.

The following examples of sentences taken from Haegeman (1991), illustrate the need for a constraint like subjacency.

(1) Poirot told me [$_{CP}$ when$_i$ [$_{IP}$ he had seen Miss Marple t_i]].

(2) Poirot told me [$_{CP}$ who$_i$ [$_{IP}$ he had seen t_i last week]].

(3) *[$_{CP}$ Who$_i$ did [$_{IP}$ Poirot tell you [$_{cp}$ when$_j$ [$_{IP}$ he had seen t_i t_j]]]]?

The subjacency condition accounts for why sentences (1) and (2) are grammatical while (3) is ungrammatical. The subjacency condition states that movement cannot cross more than one bounding node, where bounding nodes (in English) are IP and NP (Haegeman, 1991). Therefore in English, no element may be moved past more than one IP or NP. In sentences (1) and (2) above, only one bounding node (IP) has been crossed when the wh-word moves to the lower COMP, hence the sentences are grammatical. The problem with sentence (3) however is that the wh-word 'who' has crossed two IP nodes without being able to 'stop off' in the lower COMP, which is already filled. Consequently, the sentence is ungrammatical (Haegeman, 1991).

It is on the basis of ungrammatical examples such as (3) that Ross (1967, cf. Haegeman, 1991) proposed that extraction (i.e., movement of the 'who') out of a wh-phrase must be blocked. These types of sentences illustrate that when a wh-phrase is extracted from a sentential complement whose COMP is already filled by a wh-phrase, an ungrammatical sentence results.

Subjacency is a universal principle, since all languages which have movement rules must observe the subjacency condition (White, 1989). Subjacency can

accommodate variability across languages however, since what constitutes a bounding node may vary across languages. As seen above, English has IP and NP as bounding nodes. French and Italian however, have been argued to have different bounding nodes than English (Sportiche, 1981; Rizzi, 1982), namely, NP and CP rather than NP and IP

In summary, a child learning the syntax of his or her language does not have to learn the universal principles since these are innately specified What the child must learn are the specific parameter settings for different principles. In the case of subjacency, the child does not have to learn that movement cannot extend beyond more than one bounding node, (since the Subjacency Condition is a part of UG), but does have to learn what the bounding nodes for his/her language are. It is the linguistic input the child receives that triggers the parameter setting. Thus the child learning English will learn the bounding nodes of English and not those of Italian and French.

1.1.2 Second Language Acquisition

Over a billion people in the world speak more than one language fluently. In many societies, one needs to be able to speak and comprehend at least one other language in order to function fully in their community. Determining the nature of the processes underlying SLA has significant implications. The general field of SLA received a fresh impetus when Corder (1967) suggested that the processes and principles of child L1 acquisition may be parallel and indeed identical to those of adult L2 acquisition Corder suggested that obvious differences between child L1 acquisition and adult L2 acquisition[1] did not necessarily mean that the underlying processes that occur in the learning of language were inherently different. Assuming similarity, it was reasonable to determine how far child L1 acquisition processes paralleled those of learning a second language. As discussed above, the theory of UG was largely motivated to account for the phenomenon of child L1 acquisition. Therefore, with regards to SLA, an obvious but

[1] Some of those differences are the contrasts in degrees of successful acquisition Normal children always succeed in becoming fluent but many adults have difficulty reaching native-like fluency in their L2

indirect consequence of Corder's (1967) suggestions was to attempt to apply UG to adult SLA.

A second motivating factor for applying UG to SLA is that just as there is a logical problem for child L1 acquisition, so too there seems to be a corresponding logical problem for adult SLA (White, 1989). Like child L1 acquisition, the input adult L2 learners receive underdetermines the adult's final knowledge of the target language grammar. The quality of the input adult L2 learners receive is also, in certain respects, comparable to that received by children. If the L2 input is degenerate then one might require internalized knowledge of abstract properties to account for success in learning the L2. Correspondingly, if the input is highly simplified as in 'teacher talk' or 'foreigner talk' then there is still a gap between the input and the demonstrated ability of L2 learners to induce the abstract linguistic properties of the target grammar.

A third issue which motivates the application of UG to SLA is that of negative evidence. While it is generally accepted that most child learners do not receive much negative evidence, it is conceivable that adults learning in a classroom context could receive negative evidence. Teachers may explicitly inform their students as to what is grammatical and what is not through the course of instruction. Lack of negative evidence in adult SLA is still a potential problem however, because not all adult learners receive negative evidence. For negative evidence to be useful, it must be consistently available and used by the L2 learners (White, 1989). Therefore, all successful L2 learners who don't receive any negative evidence must be accounted for. Furthermore, there is no guarantee that L2 learners will heed the negative evidence that they might receive and the errors that adults typically make do not involve violations of principles of UG. It would be difficult, therefore, for the L2 learner to be corrected for an error never made (White, 1989).

Universal grammar and its principles offer interesting implications for adult L2 learners. If principles such as subjacency are innately specified, then one can investigate

whether an adult L2 learner is sensitive to this knowledge in their L2, even if their L1 has a different setting for subjacency. For example, if UG were accessible to adult language learners, then the Japanese learner of English would have knowledge of the principle of subjacency even though it was not instantiated in her L1, which lacks the relevant kind of movement. If one were able to demonstrate that adult L2 learners had access to principles of UG, then one is potentially closer to a comprehensive understanding of the unconscious knowledge accessible to an adult L2 learner.

The evidence concerning UG's accessibility or role in adult SLA has been largely equivocal (Bley-Vroman et al., 1988; Clahsen & Muysken, 1986; Ritchie, 1978, Schachter, 1989; White, 1988; Finer, 1991). Essentially, these studies attempted to ascertain whether or not UG plays a role in adult SLA, and if so, what the nature of that role might be. On at least one level of analysis then, the research has a common agenda However, the experiments carried out to date rarely have a common conclusion.

Bley-Vroman et al. (1988) investigated whether L2 learners of English, whose L1 was Korean, were able to correctly judge English sentences, the crucial, target sentences violating subjacency in English. The important aspect here is that in Korean, there is no syntactic movement and therefore, subjacency does not apply. The investigation, then, examined whether subjects would exhibit knowledge of subjacency, which they could not have acquired through L1 transfer. The results of the GJ task revealed that L2 learners were less accurate in correctly judging sentences than native speakers, but their accuracy scores were significantly above chance. The conclusions were that adults do have access to knowledge of UG; however, the authors were left pondering why their L2 learners should have lower accuracy scores than their native speakers

Ritchie (1978) also investigated the acquisition of English by adult, native speakers of Japanese. Japanese has no rightward movement rules, while English has the Right Roof Constraint which restricts where a syntactic element may be moved to the right (Ritchie, 1978). The task for the L2 learners was a GJ task where they were asked

to evaluate pairs of sentences and determine which sentence of a pair was more or less grammatical than the other The results indicate that the adult Japanese learners of English were following the constraints laid down by the Right Roof Constraint. Ritchie (1978) concludes that UG is operating in L2 acquisition. However, the subjects tested were not true adult learners (since they had exposure between the ages of 8 and 13 years).

Clahsen and Muysken (1986) reviewed previous longitudinal and cross-sectional research on the acquisition of German word order in adult L2 learning. They suggest that adult learners of German initially adopt an incorrect rule for word order in German, and then, in order to produce the proper order, apply 'unnatural' rules to generate the correct word order for German. The conclusions reached were that adult L2 learners acquire their L2 through general learning strategies and processing constraints. Furthermore, the rules that adult L2 learners hypothesize are unnatural and, therefore, do not observe the constraints imposed by UG. There are a number of problems involved with the Clahsen and Muysken (1986) paper, most crucial of which is they way the data are interpreted, i.e., one way for L1 data and another way for L2 data (White, 1989). Furthermore, their conclusions are based wholly on spontaneous 'naturalistic' production data which makes it impossible to make any conclusions about L2 competence based solely on this kind of L2 performance.

Occasionally the adult SLA studies within the UG framework have been cross-sectional and/or longitudinal and/or investigate acquisition through means of a production-elicitation procedure (Clahsen & Muysken, 1986; Flynn, 1984). Typically however, the research is more experimental in that subjects from a variety of language groups are tested and compared by means of a variety of tests which attempt to isolate specific linguistic structures as directed by the agenda of the research and which do not rely wholly on production data (Ritchie, 1978; Schmidt, 1980; Mazurkewich, 1984; Adjémian & Liceras, 1984; White, 1985, 1986; Bardovi-Harlig, 1986; Bley-Vroman et al., 1988). While some researchers feel their evidence supports the notion that principles

of UG (like subjacency) do not operate in SLA (Schachter 1989), others feel their research clearly demonstrates support for the notion that interlanguages[2] adhere to universal constraints and that specific principles can be argued to be operating in the adult L2 learner's interlanguage grammar (Schmidt, 1980, Ritchie, 1978, White, 1989)

One is left with a residual impression of indefiniteness after examining the research to date. Despite the efforts of the researchers who have contributed to the investigation of UG in SLA, a definite sense of ambiguity remains. Though researchers may obtain the results they expected, (for example, White, 1988, Bley Vroman et al., 1989; Ritchie 1978) they often are left with enigmatic results and/or further problems to face. One must examine the research more precisely in order to isolate possible variables which could contribute to the opposing findings

Schachter (1989) investigated whether knowledge of subjacency violations would be evident in subjects whose native language did not have wh-movement (and therefore, did not have subjacency substantiated in their L1). She tested subjects who were native speakers of Korean, Chinese and Indonesian learning English. Neither Korean nor Chinese has wh-movement[3]. Korean speakers are assumed not to have had any exposure to subjacency in their L1 while Chinese speakers are assumed not to have had L1 exposure to subjacency in conditions involving wh-movement. Indonesian does have wh-movement but it is more restrictive than in English.[4] The task for the subjects was to judge the grammaticality of different sentence types, either grammatical (the syntax test) or ungrammatical (the subjacency test). If these L2 learners can access UG, they should pass both the syntax and the subjacency tests. Should a subject fail the subjacency test but pass the syntax test, the data are assumed to indicate that the subjects

---

[2] Interlanguages are the L2 learner's grammar of the target language, assuming that the grammar is not yet identical to a native speaker's grammar

[3] Though Chinese does have other forms of movement, it does not have surface movement of wh elements (Schachter, 1989)

[4] The wh-word must be moved to subject position in its own clause prior to being moved to the beginning of the matrix (Schachter, 1989).

do not have access to the subjacency condition in their L2. The results of Schachter's investigation revealed that two-thirds of the L2 learners passed the syntax test, but failed the subjacency test The Korean subjects deviated the most from the native speaker norms, but were not significantly different from the Chinese or Indonesian subject groups (Schachter, 1989) The conclusions drawn from this investigation were that an L2 learner has knowledge of universal principles in their L2 but only to the extent that those principles are instantiated in their native language.

Another investigation examining subjacency's possible accessibility in L2 learning was conducted by White (1988). White (1988) tested whether native speakers of French acquire knowledge about the bounding status of IP in English. As discussed previously, it has been argued that IP is not a bounding node in French (Sportiche, 1981) while it is in English. Therefore, any movement which crosses more than one IP node in English will be ungrammatical yet grammatical in French as in (4):

(4) *To whom did Michael wonder what Janet had mailed?

A native speaker control group and two groups of intermediate level adult learners of English (whose L1 was French) were tested on various grammaticality judgment (GJ) tasks which included violations of subjacency. Both groups of adult L2 subjects achieved a high level of accuracy judging sentences with subjacency violations resulting from extraction from complex noun phrases; violations that were ungrammatical in both French and English However, only the high-level intermediate group of L2 learners reached a high level of accuracy on the wh-island violations; sentences which were ungrammatical in English, but allegedly grammatical in French. White (1988) argued that the low-level group's low accuracy is not evidence to refute UG's availability to L2 learners Rather, she offered the following hypothesis: L2 learners had to learn to reset their parameters regarding the bounding status of IP. The high level group had already learned to reset their parameters to include IP as a bounding node in English and thus were accurate in detecting wh-island violations in English. The low level group however,

initially assume that English is like French such that IP is not a bounding node and consequently erroneously evaluate wh-island violations as being grammatical (White, 1988).

2.0 General Methodological Problems

These two studies (White, 1988 and Schachter, 1989), illustrated the conflicting nature of much of the research to date. Schachter (1989) concluded that sensitivity to violations of subjacency is only possible if the subjacency condition is part of the L1. White (1988) however, suggested that even when the subjacency condition is a part of the native language, L2 learners may initially be unaware of subjacency violations in their second language. White (1988) assumed that adult L2 learners can reset parameters when learning an L2 with a different setting. However, Schachter (1989) assumed that any knowledge of subjacency an adult L2 learner will have arisen from transfer from their L1. Incompatible assumptions may be better understood by examining differences in methodology.

The Schachter (1989) investigation for example, suffered from a number of methodological flaws. An obvious flaw was that it was impossible to be sure that the subjects' responses were not reflecting a response bias. If her subjects had a tendency to accept all sentences, then they would have passed the syntax test (by accepting them all) and failed the subjacency test (by accepting them also). Schachter (1989) reported that this is exactly what her subjects did but considered her results to be evidence against UG's availability in SLA. Furthermore, Schachter (1989) did not present grammatical wh-questions to the subjects. One cannot be sure, therefore, that her subjects had knowledge of wh-questions at all, regardless of whether or not they violated subjacency. Additionally, Schachter (1989) did not pre-test her stimuli to ensure that her sentences were appropriate. The native speaker controls had significant difficulty in judging the grammaticality of the noun complement sentences. Correspondingly, it is on this

structure that the non-natives perform the most poorly. One cannot rely on the data therefore, since even the native speakers were unable to yield reliable judgments.

White's (1988) investigation also suffered from methodological problems. One of the key comparisons made was between the two groups of experimental subjects; the low versus the high intermediate groups. These subjects were categorized on the basis of teacher ratings. However, the same teacher did not rate all of the subjects. The two groups were recruited from two separate institutions. One cannot be sure therefore, that the different teacher ratings were comparable, and consequently whether the two groups themselves were comparable. Any marginally anomalous results could be attributable to different levels of proficiency in different groups. White (1988) had no way of determining therefore, the true level of proficiency of her subjects other than a cloze test. A more standard test of proficiency in addition to the cloze test she administered may have shed some light on the puzzling result that her two groups performed differently on the wh-island structures but identically on the cloze test.

It is clear that a neglect of methodological issues can and does detrimentally affect the explanations of results of research, regardless of the domain within which it is conducted. There are a number of methodological considerations researchers should notice when investigating second language acquisition, some of which are outlined in Chaudron (1983) and Birdsong (1989). Some experimental issues which have been discussed are; response biases, the validity of experimental measures, background linguistic variables of subjects, pre-test measures of subjects, explicit instructions and convergent and validating evidence. However, one additional issue that has not been disucssed in previous research and which has only recently received attention concerns modality

2.1 Modality

Not all of humans' sensory systems are utilized equally. Vision, for example, is dominant in human beings (Rock & Victor, 1964). The priority of vision over touch was

demonstrated in a task where a square was made to look like a rectangle whose sides appeared in the proportion of two to one. When subjects both felt and saw the square through a distorting lens, (which produced an optical compression of width) they perceived the square as a rectangle. The stimulus therefore, was perceived on the basis of the distorted visual input rather than the undistorted tactual one (Rock & Victor, 1964)

Research on modality differences suggests that verbal material presented aurally and visually is processed in different parts of the memory system and by different mechanisms (see Penney, 1989 for a comprehensive review). The general view is characterized in the separate streams hypothesis (Penney, 1980) which proposes that subjects are able to process information in the visual and auditory modalities independently and without interference (Rollins & Hendricks, 1980)

Representative evidence to support the separate streams hypothesis comes from research in which subjects have to shadow input from one modality while monitoring input from another. (Dennis, 1977; Shaffer, 1975) For example, Dennis (1977) required subjects to listen to an auditorily presented message while at the same time, monitoring either an auditory or a visual list of words. There were more errors on a detection task when the word list was presented auditorily than when it was presented visually Therefore, when subjects were given additional material to monitor in a different modality, there were differences in the accuracy at detecting target items Subjects also demonstrate superior performance when stimuli are presented in two modalities rather than one. For example, in Frick (1984), there was a dual mode condition where four numbers were presented visually and auditorily to the subjects. There was also a pure visual and a pure auditory condition. Recall of items was higher when two presentation modalities (dual mode condition) were used instead of one (Frick, 1984) Auditory (Broadbent, Vines & Broadbent, 1978) and visual (Penney, 1974) selective interference effects, where distractor items are presented in a different modality between presentation of target items and recall, also supports the separate streams hypothesis. When the

distractor items are presented in the same modality, recall is detrimentally affected. There is no adverse affect, however, when the distractors are presented in the alternate modality (Penney, 1989). Furthermore, there is evidence that information is organized according to the modality of presentation. Subjects seem to prefer to recall stimuli according to the modality of presentation when there are different ways the stimuli are presented. For example, when stimuli were presented in two modalities, in two languages, and according to different semantic categories, subjects recalled the items according to the modality of presentation and not the language of presentation or to the semantic category (Ronnberg, Nilsson and Ohlsson, 1982). A final line of evidence which supports the notion that information is processed and stored differently depending on the modality of presentation comes from the observation that there are modality-specific deficits in short-term memory (Shallice & Warrington, 1977).

2.2 Modality in Second Language Acquisition Research

While modality of stimulus presentation is a methodological issue for SLA research which has largely been neglected, some researchers have recently begun to examine modality's effect on subjects' performance on metalinguistic tasks.

Johnson (1992), conducted a follow-up study to one which had previously demonstrated a critical period effect with second language learners (Johnson & Newport, 1989). The original investigation offered evidence to indicate that children have an advantage over adults in acquiring a second language. The Johnson and Newport (1989) study examined the hypothesis originally proposed by Lenneberg (1967), that language could not be successfully acquired after puberty. The proficiency of native speakers of Korean and Chinese who had arrived in the U.S. between the ages of 3 and 26 years was tested on a GJ task, presented aurally, which included sentences highlighting a variety of aspects of English grammar (e.g., past tense, plural, third person singular etc.) The results indicated that there was a negative linear relationship between age of exposure and proficiency such that the older the subjects were when they arrived, the less proficient

around the age of 16, however, performance on the GJ task was lower and there was no systematic relationship between age of exposure and proficiency in the language (Johnson & Newport, 1989).

The Johnson (1992) investigation, which also focused on language-specific rules, tested the same subjects and used the same stimuli as in the earlier Johnson and Newport (1989) experiment. The main motivation for conducting the Johnson (1992) study was to determine whether the late L2 learners in the Johnson and Newport (1989) investigation exhibited poor performance because of artifacts of the auditory task. Subsequently, the only differences between the two investigations were that fewer of the subjects participated and that the stimuli were presented in written versus auditory form. Johnson (1992) identified three possible ways an auditory task could detrimentally affect the outcome of the subjects' judgments: 1) that information must be processed on-line in an auditory task 2) that stimuli must be phonologically decoded and any difficulties the subjects may have with the L2 phonology will cause difficulty, and finally that 3) the auditory task in the original study was timed such that subjects could not control how much time they were allowed to take in processing the sentences. The subjects in the Johnson (1992) study were asked to complete a GJ task which contained the same materials as the Johnson and Newport (1989) investigation. The crucial difference, however, was that the stimulus materials were presented visually, on paper, and the subjects were instructed they could take as much time as they pleased to reach their decision with each sentence. The results indicated that there was still a strong negative correlation between age of exposure and proficiency in the L1. However, performance on the GJ task was markedly improved on the written, as compared to the aural version of the task. The adult learners made more than twice as many errors in the auditory version than the visual version of the GJ task. The visual task therefore, seemed to be less discriminating than the auditory task (Johnson, 1992).

Johnson (1992) offers two possibilities as to why performance was significantly improved from the auditory to the visual task. The first suggestion is that the performance on the auditory task is worse than what would be predicted from the subjects' grammatical knowledge. In other words, difficulties inherent in the task mask the subjects' true grammatical competence. Placing subjects in a position where they have to process information on-line and where they are unable to reflect meta-linguistically on the task items interferes with their abilities to perform. The second suggestion is that a written GJ task somehow inflates grammatical competence such that the performance on the task is an augmented reflection of what the subjects' true grammatical competence actually is. Formal language training, for example, may enhance their metalinguistic skills such that when on-line burdens are eliminated, these skills become accessible. Johnson (1992) acknowledges that while these two possibilities are not mutually exclusive, her data cannot discriminate between them.

Additional evidence to demonstrate a modality effect in a GJ task was illustrated in Haig (1991). Haig (1991) tested native Chinese speakers who learned English as adults. Half of these subjects were given a written task of judging sentences, some of which violated subjacency, and the other half were assigned to an aural task, in which they had to judge sentences on the basis of an aural presentation. Subjects in the aural task had lower accuracy scores in correctly rejecting subjacency violations than subjects in the written task. Thus, modality differences can be found in tasks which investigate adult L2 learners' knowledge of universal constraints. Subjects who received a written version of a grammaticality judgment (GJ) task were significantly more accurate in rejecting subjacency violations than those who received the GJ task aurally.

Other research which investigates the status of subjacency in adult L2 acquisition may find equally significant differences between task types. Johnson and Newport (1991) for example, offered evidence which allegedly established that subjacency does not differ from language-specific structures with regards to critical period effects.

Universal properties are claimed to also undergo broad deterioration as learners become increasingly mature (Johnson & Newport, 1991). In the Johnson and Newport (1991) investigation, all the stimuli were presented aurally. Based on the work by Johnson (1992) and Haig (1991) one might predict that these subjects would have performed significantly more accurately if the stimuli had been presented in written form.

The studies reported above suggest that subjects will respond differently on a GJ task when it is presented aurally as opposed to visually. Up until this point however, the importance of these processing differences have not been considered sufficiently. Rather substantial conclusions are made on the basis of results taken from only one modality (e.g., Johnson & Newport, 1991). Furthermore, much research conducted on adult SLA has used GJ tasks. There is a great need, therefore, for further empirical evidence to illustrate that modality of stimuli presentation is an important methodological concern.

The present investigation was designed to provide further empirical evidence regarding the issue of modality of presentation of a GJ task for adult L2 learners. The experiment addresses the issue of modality by presenting subjects with sentences to judge on a computer, where half of the subjects received the sentences visually on the computer screen and the other half heard the same sentences over headphones. The sentences in both modalities were presented for exactly the same amount of time. Presentation time is an important issue. To illustrate, Johnson (1992) found significant differences for modality of stimulus presentation, however, in the aural version of the task presentation time was dictated by the experimenter. In the written version, Johnson's (1992) subjects were encouraged to take as much time as they needed. A crucial aspect of the Johnson (1992) experiment was to compare it to the Johnson and Newport (1989) study. However, it is difficult to make such a comparison based on such disparate presentation times. This methodological problem was alleviated in the present experiment. For example, if it took 4000 msecs. to present a sentence aurally, then that same sentence was presented visually for 4000 msecs.

Both target and distractor sentences were presented to the subjects. The target sentences are those which violate subjacency in English but, according to Sportiche (1981), would be grammatical in French. There were also grammatical target sentences in which embedded wh-questions was presented to the subjects. The inclusion of these sentences was important to ensure that the subjects at least were able to process declarative versions of the wh-island structure. The distractor sentences were syntactically complex sentences which had the same number of words as the target sentences. Half of the distractors were grammatical and the other half were ungrammatical.

The present experiment is more highly controlled methodologically than typical experiments in the UG/L2 paradigm. All of the sentences to be judged were constructed such that each had exactly the same number of words. Many experiments using a GJ task have neglected to ensure that each sentence is as similar to each other as possible. Some sentences may be very short, and syntactically simple, while others (especially ones included to test knowledge of subjacency) will be long and potentially difficult to parse. It is important therefore, not to bias the subjects into recognizing which items are target items by occasionally presenting long and complex sentences.

Additionally, the overall number of sentences to judge were restrained to twenty. Johnson (1992), for example, presented 276 sentences to her subjects. With such a large number of stimuli, one could always argue that any poor performance effects could be due to subject fatigue. A GJ task can be difficult to attend to under the best of circumstances. Therefore, it is advantageous not to fatigue the subjects too much by providing them with an extravagant number of sentences to judge.

A further control in constructing the sentences involved the specific words chosen. There were no proper nouns used in the sentences to ensure that there would be no possible influence or bias of the subjects in reacting to certain names. Furthermore, the nouns in the sentences were all high frequency nouns as identified by Kucera and

Francis (1967). Such a control reduces the likelihood that subjects will find a sentence difficult because they could not understand what a specific lexical item meant.

An additional problem encountered in the literature, which is addressed in the present experiment, regards the proficiency of the subjects. Many experiments do not report standardized measures of proficiency in the target language. Obviously such a measure is necessary to ensure a) homogeneity of the subjects and b) that subjects have at least the necessary proficiency to be able to perform the task. The adult L2 learners of English also had to complete the standard TOEFL test[5] (in addition to a cloze test) to obtain a standardized measure of their proficiency in English.

Native French speakers who learned English as adults should be aware of the status of IP as a bounding node in English if knowledge of the subjacency condition is available to adults. Furthermore, if the Haig (1991) and Johnson (1992) results are reliable, then subjects who receive stimuli visually should be faster and more accurate at correctly rejecting subjacency violations than subjects who receive the stimuli aurally

3.0 Method

3.1 Subjects

Eighty subjects participated in the experiment. These subjects included native speakers of English (L1.E), native speakers of French (L1.F), native speakers of English who learned French as adults (FSL), and native speakers of French who learned English as adults (ESL). All subjects were undergraduate or graduate students from various faculties at McGill University in Montréal. None of the subjects were graduate or undergraduate students in linguistics. The subjects were recruited through an advertisement placed in the McGill newspaper and posted around the McGill campus. The individuals responded to the advertisements by phone, at which time the experimenter held an informal interview with the potential subjects to determine whether

[5]Test Of English as a Foreign Language.

they fit the criteria of the experiment. For the L2 learners, the criteria were that they started learning their L2 after puberty (minimum age of 13) and that they were fluent in their L2 at the time of testing.

Twenty subjects from each of the four language groups were tested. Subjects from each language group were randomly assigned into the two modality conditions; aural and visual. Thus, there were ten subjects per language group in the aural condition and ten subjects per language group in the visual condition. Overall, there were forty subjects in the aural and visual conditions respectively.

The overall mean age of the subjects was 22.9 years. The mean age of the ESL, FSL, L1.E and L1.F subjects was 22.5, 21.5, 26.5 and 21.3 respectively. The mean age of all subjects in the aural condition was 22.9 while the mean age of all the subjects in the visual condition was 23.0. For the L2 learners, the mean age at which they started learning their L2 was 15 years. The ages at which the L2 learners started learning their L2's ranged from 13 to 23 years.

Thirty of the subjects tested were males and fifty were females. The number of males in the ESL, FSL, L1.E and L1.F language groups was 10, 8, 6 and 6 respectively. The number of females in the ESL, FSL, L1.E and L1.F language groups was 9, 15, 14 and 12 respectively.

3.2 Materials

Overall, there were fifty 9-word sentences constructed to be judged in the grammaticality judgment (GJ) task. All sentences are presented in Appendix A. There were thirty experimental sentences and twenty practice sentences. Within the thirty experimental sentences, 20 were targets and 10 were distractors.

There were twenty experimental target sentences which consisted of the wh-island structure. Ten of these were ungrammatical wh-questions which violated subjacency (ungrammatical targets) such as *"What did the teacher know why the student said?". The other ten were corresponding declarative grammatical embedded wh-questions

(grammatical targets), such as "The parents discovered why the children played the music". The nouns used in the target sentences ranged in frequency from 57 to 1207 Thus, all nouns were high frequency according to Kucera and Francis (1967).

There were two different versions of the experiment with regards to the grammaticality of the target sentences. For half of the subjects (group A), the subjacency violations they received were converted into the grammatical declarative targets the other half of the subjects received (group B) Correspondingly, the grammatical targets in group B were converted into the ungrammatical targets in group A. For example, group A received the sentence *"What did the teacher know why the student said?". Group B on the other hand, received that sentence as a grammatical target, namely, "The teacher knew why the young student said hello". At the same time, group A received the declarative sentence "The king understood why the enemy revealed the plan" while group B received that declarative grammatical target as a violation of subjacency, namely; *"What did the king understand why the enemy revealed?" Thus, for each declarative grammatical target group A received, group B received those same sentences as subjacency violations, while the subjacency violations group A received were the declarative grammatical targets group B received. The presentation order of the sentences was counterbalanced such that the first half of the sentences presented to one group of subjects was the last half of sentences received by the other group Thus, the order of stimulus presentation and the grammaticality of target sentences were counterbalanced.

The remaining experimental sentences were distractor sentences. The distractor items were constructed to control for certain possible effects which may have arisen if they were not included. All the distractor sentences contained embedded sentences so they resembled the target sentences in syntactic complexity. There were four types of distractor sentences; two declarative-grammatical sentences ("The couple spoke with the person who lives downstairs"), three declarative-ungrammatical (*"The judge went to the

store the surprised manager"), three question-grammatical ("What did the parents think the young children wanted?") and two question-ungrammatical (*"When the analyst did remember the appointment was?") It was necessary to construct the four kinds of distractor sentences to control for possible effects of responding to the targets. The target sentences were grammatical declaratives and ungrammatical wh-questions. Therefore, it was necessary to have both ungrammatical and grammatical declaratives, and ungrammatical and grammatical wh-questions to ensure that, if there were differences in responding to the target sentences, it was not due to grammaticality or ungrammaticality in general, or general differences in processing declarative versus wh-question sentences.

Twelve practice trials were also created. These were sentences which also contained nine words and were syntactically complex. Half of these were grammatical, the other half ungrammatical, and half were questions and the other half declaratives. There were also eight dummy trials included within the experimental trials. The first five and the last 3 sentences that each subject was presented, within the experimental session, were dummy trials.

Once the stimulus materials were assembled, they were translated into French by a francophone syntactician from the Department of Linguistics at McGill University in Montréal. Thus, there were two sets of 50 sentences, one in English which were presented to the L1.E and ESL subjects, and the French translations thereof which were presented to the L1.F and FSL subjects.

The sentences were presented on a computer (Macintosh SE) running Psychlab (v.0.85) (Gum & Bub, 1988). The aural sentences were digitally recorded onto a computer (Quadra 700) and manipulated using a soundeditor (Soundwave, 1.1). All subjects used earphones. A keyboard was used to record their responses.

The TOEFL[6] test (ETS, 1993) was used to measure the ESL subjects' proficiency in English. A cloze test was used to measure all subjects' proficiency levels in their second or native languages.

3.3 Pilot Testing

Once the stimuli had been constructed and translated, they were pilot tested on native speakers of English and French, respectively, to ensure that native speakers would judge the sentences the way the experimenter meant them to be judged. This was an important step, since if the native speakers were not "getting it right" then any results from the L2 learners would be meaningless. Both the English sentences and the French translations were assembled in a questionnaire format where a sentence was presented, centered on one page. Directly underneath each sentence was a choice of 'good' or 'bad'. A Likert type scale was also included with each sentence. The scale ranged from 1 to 5, where 1 indicated complete certainty in the judgment and 5 indicated complete uncertainty regarding the judgment. The questionnaires for both the English and French sentences were given to 10 native speakers of English and 10 native speakers of French (who were not linguists or linguistics students). These native speakers were asked to rate the grammaticality of each sentence as good or bad, and also to indicate how certain they were of their judgment on the scale of 1 to 5. These native speakers were also asked not to reconsider a sentence once having made a judgment. There was only one sentence per page to facilitate the subjects conforming to this instruction. The pilot testing ensured that all the sentences which were supposed to be judged as grammatical or ungrammatical, as prescribed by linguistic theory, were so judged by native speakers.

3.4 Procedure

Once the pilot testing was completed, a male native speaker of English and a male native speaker of French recorded the fifty English and French sentences respectively.

---

[6]Test of English as a Foreign Language

Each recording was then measured to determine the sentence's length in milliseconds. This measurement was used so that presentation time of each sentence across modalities was identical. That is, for a given sentence in the aural condition, that same sentence was presented in the visual condition for exactly the same amount of time.

The experimental session for the ESL subjects proceeded as follows. Each ESL subject completed sections 1 and 3 of TOEFL. Section 1 of TOEFL is a listening comprehension task and section 3 is a reading comprehension task. Once each ESL subject had completed the appropriate sections of TOEFL (which usually lasted approximately 1 hour), they were given a cloze test. The cloze test was a passage in English where every sixth word was removed. Subjects were instructed to fill in the blanks of the passage with the appropriate English word such that the passage would be grammatical and coherent. After completing the cloze test, the ESL subjects then participated in the experimental manipulation. The experimental session for the FSL, L1.E and L1.F subjects was identical. However, they did not have to complete the TOEFL test. Their cloze test was the same passage in French where every sixth word was removed.

After completion of the cloze test for all subjects, the experimenter gave the subjects the instructions for the experiment. All subjects were required to put on the earphones when participating in the experiment. They were told that they would be hearing or seeing sentences in English (or French). They were informed that at the beginning of each sentence, there would be a tone and a string of asterisks (which served as a fixation point in the visual condition) which would appear on the screen together to prompt the subjects that a sentence was about to be presented. The subjects were requested to listen to (or read) the sentence carefully. There were also advised that after the sentence was presented, they would hear a very short low beep which would indicate to them that they should respond. If they felt that the previously presented sentence was something a native speaker of English (or French) might say, they were asked to press a

key labeled 'G' (for Good) on the keyboard as quickly as possible after the short low beep. If they felt however, that it would be impossible for a native speaker to ever utter such a sentence, they were asked to press a key labeled 'B' (for Bad) on the keyboard as quickly as possible after the short low beep[7].

The subjects were expressly requested only to judge the sentences as bad if there were no conceivable conditions under which a native speaker might produce an utterance like the sentence presented. The subjects were specifically informed that even if they personally would never produce such a sentence, but they knew it was acceptable in English or French, they were not to judge the sentence as bad. Furthermore, they were instructed not to judge a sentence as bad if they felt that the words within the sentences were not ones they normally used, but knew that they were appropriate words for the sentence.

At the beginning of each trial, a string of asterisks appeared centered on the computer screen. At the same time a tone was heard over the earphones at middle C. Both the middle C tone and the asterisks were presented for 500 msecs. After an interval of 100 msecs., a sentence was presented. In the audio condition, the recordings were presented. In the visual condition, however, a written version of the auditory sentence was presented for exactly the same amount of time on the computer screen. Sentences in the visual condition were presented in the font Helvetica, at size 14. Immediately upon completion of the presentation of each sentence (in both modalities), a low beep, one octave below middle C, was presented for 20 msecs. The computer recorded the subjects' responses and the latency between the low beep at the end of each sentence presentation and when the subjects pressed a button to make their grammaticality judgment.

The subjects were given the twelve practice trials first and after they completed these, were given an opportunity to ask any questions. When the subjects felt ready, they

---

[7]The 'Good' key was L on the keyboard and the 'Bad' key was S.

were given the 28 experimental trials (the first 5 and last 3 being dummy trials). Upon completion of the experiment, the subjects were de-briefed as to the nature and purpose of the experiment in which they had just participated.

4.0 Results

A discussion of the subjects' performance on the proficiency tests will be the starting point for the overall discussion of the results. The analysis of the subjects' accuracy at judging the grammaticality of the sentences will then be discussed. Finally, the analysis of the data of the reaction time (RT) to judge the sentences will be addressed.

As described above, two experimental orders of the sentences were used to eliminate any possibility of effects due to the presentation order. Furthermore, in order to eliminate any effect of the sentence items themselves, each ungrammatical wh-target had a grammatical counterpart. Half of the subjects thus received grammatical versions of sentences which were ungrammatical for the other half of the subjects (and vice versa). Initial analyses of variance indicated that neither order, nor version had any effect on grammaticality judgment nor on reaction time. These two variables were therefore omitted from any further analysis.

4.1 Proficiency Tests

The proficiency of the ESL subjects in English was measured by using the TOEFL (TOEFL, 1993). All the ESL subjects achieved high scores on the TOEFL test. The results of the ESL subjects' scores on the TOEFL test are presented in Table 1. Seventeen of the twenty subjects had a score above 600 which is the highest range identified by TOEFL and is considered to indicate proficiency in English. Three subjects scored between 550 and 599. Scores within this range are considered to indicate enough proficiency in English for the candidate to attend an English language institution (TOEFL, 1993).

Table 1

ESL Subjects' scores on the Test of English as a Foreign Language (TOEFL)

Subject No.	TOEFL Score
1	630
2	650
3	585
4	660
5	635
6	625
7	660
8	650
9	585
10	605
11	600
12	610
13	640
14	610
15	570
16	655
17	675
18	610
19	640
20	660

Lowest score = 570
Highest score = 600

The cloze tests were scored according to an exact matches procedure, in which only exact matches to the original paragraph were considered correct responses. The results of the cloze tests for the four language groups are presented in Table 2. An analysis of variance was applied to the data which revealed a significant difference between the four language groups on the cloze test where [$F(3,76) = 23.49, p < 0.00$]. A Scheffé test indicated that the ESL group accuracy was significantly different from the FSL group accuracy but was not different from the L1.E group accuracy. This difference in accuracy would suggest that the ESL subjects' L2 proficiency was higher than the FSL group, since the passages were identical with the exception that one was in English and the other in French. The FSL group accuracy mean percent correct was 45 while the ESL's group accuracy mean percent correct was 57. The ESL subjects were highly proficient since, as the analysis reveals, they were not significantly different from the

L1.E subjects. The FSL subjects also had a lower mean percent correct than the L1.F subjects. There was no difference on the cloze tests between the two native speaker groups (L1.E and L1.F).

Table 2.

Percent correct on cloze tests for the four language groups (ESL, FSL, L1.E, L1.F).

Language Group	% Correct
ESL	57
L1.E	58
FSL	45
L1.F	67

4.2 Accuracy Results

The design of the experiment adheres to a mixed design with three between-subject factors: 'Modality' (aural vs. visual), 'Native' (native vs. L2 learners) and 'Language' (English vs. French sentences); and two within-subject factors: 'Target' (target vs. distractor sentences), and 'Grammaticality' (grammatical vs. ungrammatical sentences). All analyses applied a multivariate repeated measures analysis of variance to the data using SYSTAT v. 5.2.1 (Wilkinson, Hill & Vang, 1993) Multivariate General Linear Hypothesis; General Linear Model. A complete source table outlining all of the accuracy results is presented in Appendix B. The means and standard deviations for the different language groups' accuracy on the target sentences for both modalities is presented in Table 3.

Table 3.

Mean number correct on the grammatical and ungrammatical targets in the aural and visual condition.

	AURAL				VISUAL			
	Grammatical Target		Ungrammatical Target		Grammatical Target		Ungrammatical Target	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ESL	4.5	223	3.2	.359	4.9	223	4.3	359
FSL	4.5	.223	3.8	.359	4.3	223	3.9	359
L1.E	5.0	.223	4.1	359	4.8	223	4.7	.359
L1.F	4.7	.223	4.8	.359	4.7	223	4.3	359

The main questions addressed by the experiment concern modality effects overall, and specifically, in those sentences which test UG's availability to L2 learners Therefore, only the results which address these questions directly will be discussed here

The accuracy results relevant to modality are presented in Figures 1 through 4 With reference to the accuracy data, the analysis reveals that there was no main effect for modality. This result is presented in Figure 1, which illustrates all subjects' overall accuracy in judging all sentences in the aural and visual modality

Figure 1. The effect of modality on accuracy in grammaticality judgment.

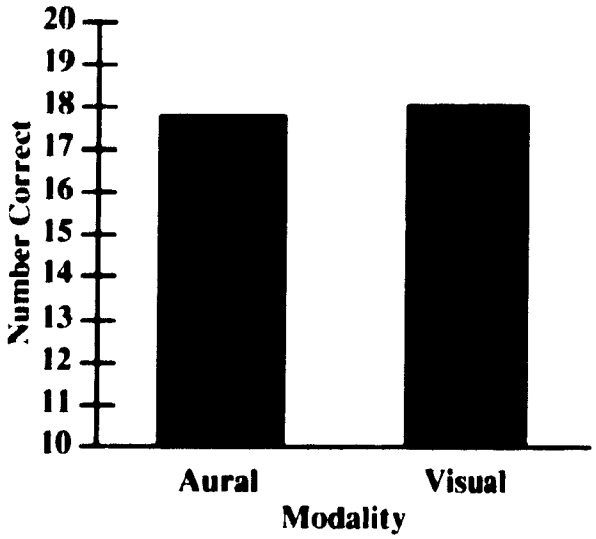
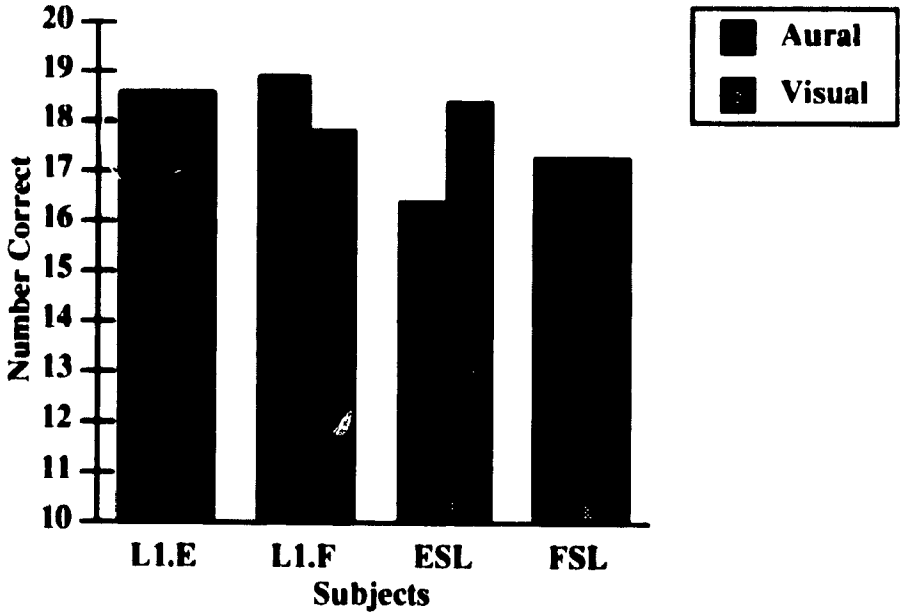


Figure 2 represents how the different language groups responded overall in the different modalities. This interaction is not significant.

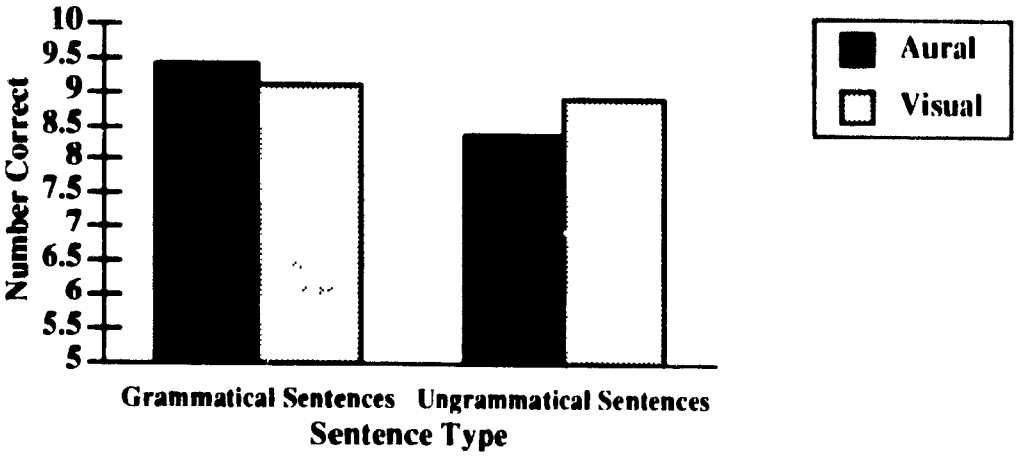
Figure 2. The effect of modality for the four language groups.



There was an effect of modality which interacted with the grammaticality of the sentences. Subjects were less accurate at judging the ungrammatical sentences in the

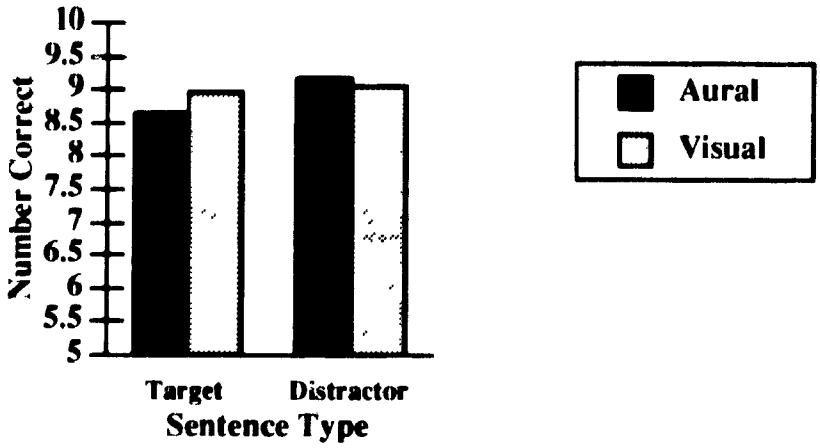
aural modality than in the visual modality. Accuracy at judging the grammatical sentences was not affected by modality. This interaction was reliable [$F(1,72) = 3.83$, $p < 0.019$] and is illustrated in Figure 3. Generally, the grammatical sentences were judged more accurately than ungrammatical ones [$F(1,72) = 7.503$, $p < 0.001$], but Figure 3 illustrates how this grammaticality effect interacts with modality. Thus, there is a clear modality effect in that processing information in an auditory task leads to reduced accuracy on ungrammatical sentences.

Figure 3. The effect of grammaticality and modality on accuracy in grammaticality judgment.



However, this modality effect was not statistically reliable when the grammatical and ungrammatical target sentences are collapsed. Figure 4 shows the means of the subjects' accuracy on the target and distractor sentences in the two modalities.

Figure 4. The effect of target and modality on accuracy in grammaticality judgment



Other aspects of the accuracy results are presented in Figures 5 through 8. The ungrammatical targets were judged significantly less accurately than the grammatical targets [$F(1,72) = 4.278$, $p < 0.002$]. As illustrated in Figure 5, subjects were less accurate at judging subjacency violations than either grammatical sentences involving embedded questions or any of the distractor sentences. As is evident from the figure, subjects' accuracy on the grammatical target sentences was quite high; therefore, it would seem that subjects were able to correctly identify the grammatical sentences involving embedded questions. When subjacency is violated, however, subjects overall were somewhat less accurate at recognizing these sentences as ungrammatical.

Figure 5. The effect of target and grammaticality on accuracy in grammaticality judgment.

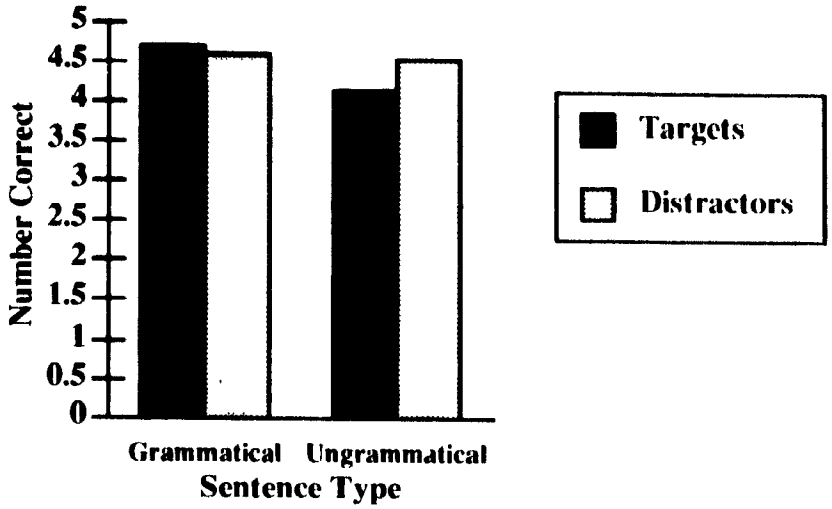
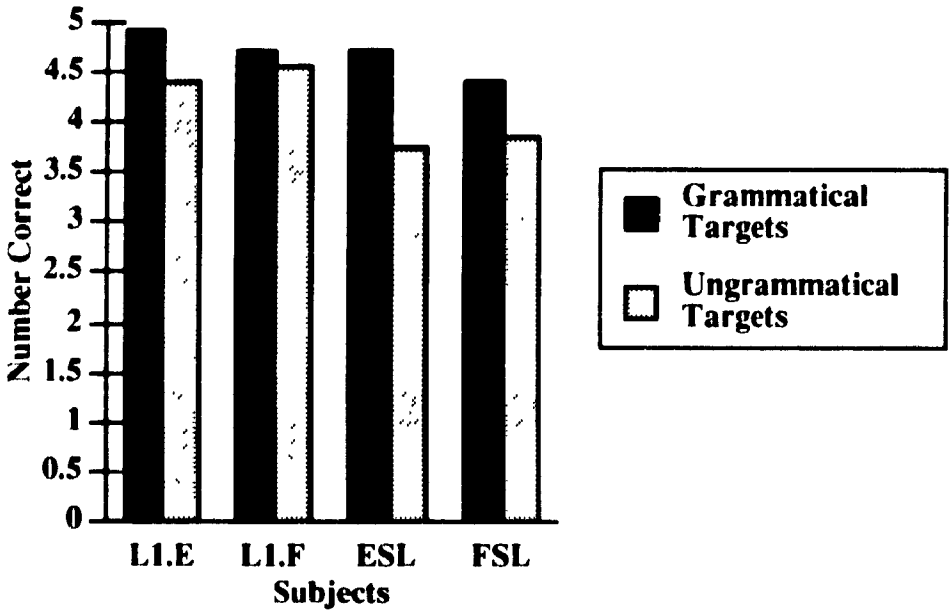


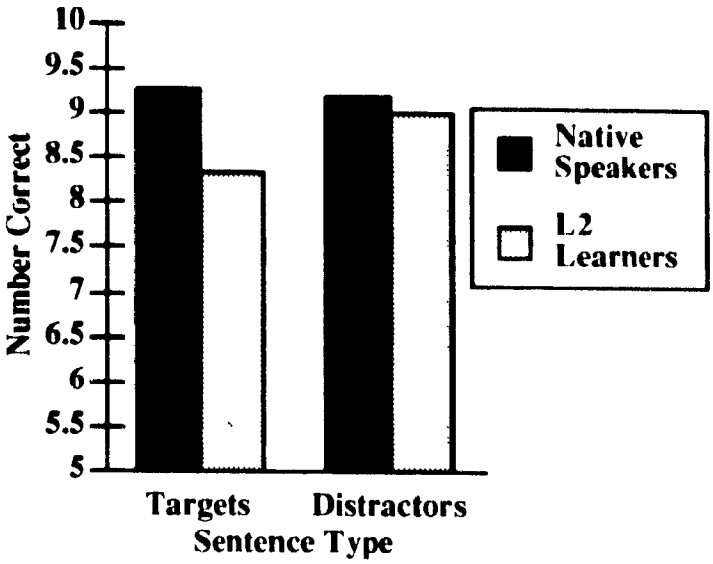
Figure 6 shows the accuracy of the subjects in each of the four language groups on the grammatical and ungrammatical target sentences collapsed across modalities. The interaction presented in Figure 6 is not statistically significant but is certainly meaningful since the analysis revealed that when the four different language groups (ESL, FSL, L1.F and L1.F) are examined separately, they were not significantly different at judging sentences which contain violations of principles of UG. All language groups (including the L2 learners) showed the same patterns of responding, suggesting that all subjects are able to respond to target sentences with equal accuracy. This result would imply that if native speakers are accessing knowledge of UG to make their judgments, it is entirely possible that L2 learners are doing so also.

Figure 6. The effect of subjects and target on accuracy in grammaticality judgment.



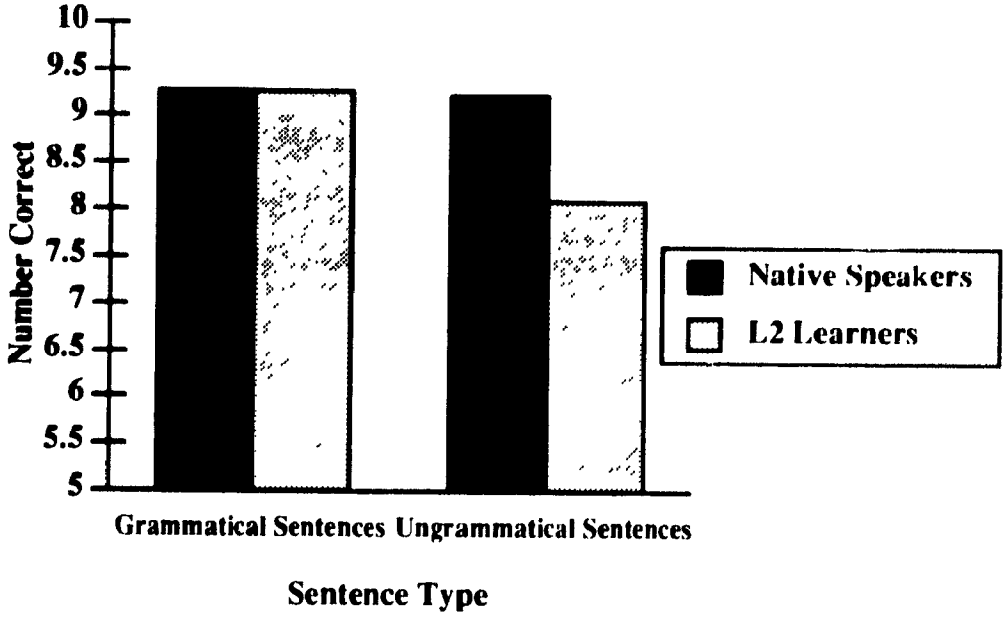
When all native speakers were compared to all L2 learners, native speakers were considerably more accurate at judging the sentences, both the wh-island structure and the distractor sentences, than the L2 learners [$F(1,72) = 2.63$, $p < 0.05$]. This interaction is illustrated in Figure 7. L2 learners were more accurate at judging the distractor sentences than the targets. While there was a substantial difference in accuracy between the native speakers and L2 learners on the target sentences, there was less of a difference in accuracy on the distractor sentences.

Figure 7. The Effect of Target and Nativeness on accuracy in



In general, native speakers and L2 learners did not differ in their accuracy of judging grammatical sentences, yet were considerably different on accuracy on ungrammatical sentences [$F(1,72) = 8.76$, $p < 0.004$] as is illustrated in Figure 8.

Figure 8. The effect of grammaticality and nativeness on accuracy in grammaticality judgment.



In summary, modality clearly interacts with grammaticality since subjects were less accurate at judging ungrammatical sentences in the aural than the visual modality. Native speakers were more accurate than L2 learners in judging all types of sentences, and the L2 learners were less accurate at judging the target sentences (both grammatical and ungrammatical) than the distractors. The sentences which violated subjacency yielded a lower accuracy score overall than the other sentences, an effect which is more pronounced for the L2 learners than the native speakers. When the four language groups are compared, there is no interaction between language of subjects (ESL, FSL, L1.E, L1.F) and grammaticality of the target sentences (subjacency violations or grammatical targets). This result suggests that L2 learners are, at minimum, performing the GJ task in a similar way to the native speakers. Correspondingly, if native speakers are assumed to be accessing knowledge of principles of UG to make their judgments, the L2 learners might also be accessing similar types of information.

4.3 Reaction Time Results

An analysis parallel to that carried out on the accuracy data was performed on the reaction time (RT) to judge the sentences. The complete source tables from the analysis of variance are presented in Appendix C. The means and standard deviations for the different language groups' reaction times on the target sentences for both modalities are presented in Table 4.

Table 4.

Mean reaction time on the grammatical and ungrammatical targets in the aural and visual condition.

	AURAL				VISUAL			
	Grammatical Target		Ungrammatical Target		Grammatical Target		Ungrammatical Target	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ESL	1627 70	190.87	1546.21	190.02	701.64	190 87	1100 44	190 02
FSL	856.67	201.20	1101 66	200 30	645.65	190 87	925 82	190 02
L1.E	534.28	190 87	798 14	190.02	486.92	190 87	736 46	190 02
L1.F	564 35	190.87	713.03	190 02	399.28	190.87	570.47	190 02

Prior to the analy irate responses were removed from the data and replaced with the mean o. ect responses for the given category within which the error was located. Thus, only ʀ Γs to correct judgments were included in the analysis. It is important to ensure that only correct judgments are included in the RT analysis, otherwise such an analysis would not be meaningful If the data were made up of both accurate and inaccurate responses, any potential speed-accuracy effect could not be controlled. Furthermore, an effect may be found in the data which may be due to the inaccuracy of the responses, but might be misunderstood to be an effect of one of the variables tested. To remove these methodological problems from affecting the RT analyses, the reaction times to inaccurate responses were removed

RT data naturally yield a skewed distribution. Therefore, in order not to violate the assumption of a normal distribution, a log transformation was applied to the RT data to yield a more normal curve and subsequently reduce the influence of the skew.

As in the accuracy analysis, the focus of the investigation was to determine whether there would be differential responding to sentences overall, and specifically to violations of subjacency in one modality than the other. The modality effects on reaction time are illustrated in Figures 9 through 12. Unlike the accuracy analysis, there was a reliable main effect of modality, where subjects in the aural modality were slower to judge sentences overall than the subjects in the visual modality, [$F(1,71) = 4.324$, $p < 0.041$]. This effect is illustrated in Figure 9.

Figure 9. Main effect of modality on reaction time to grammaticality judgment.

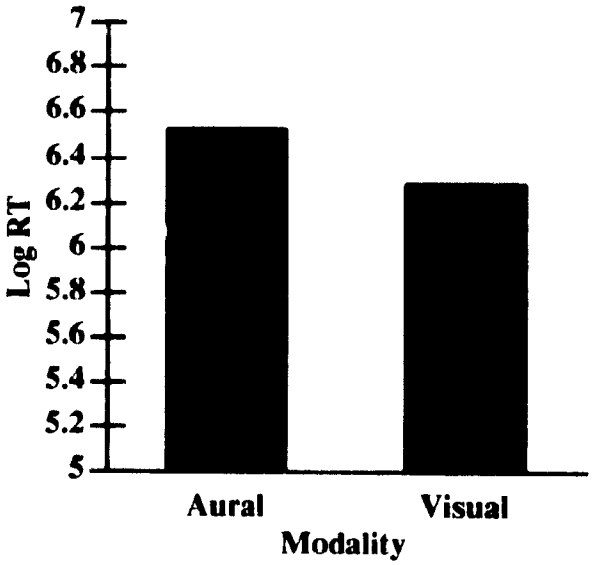
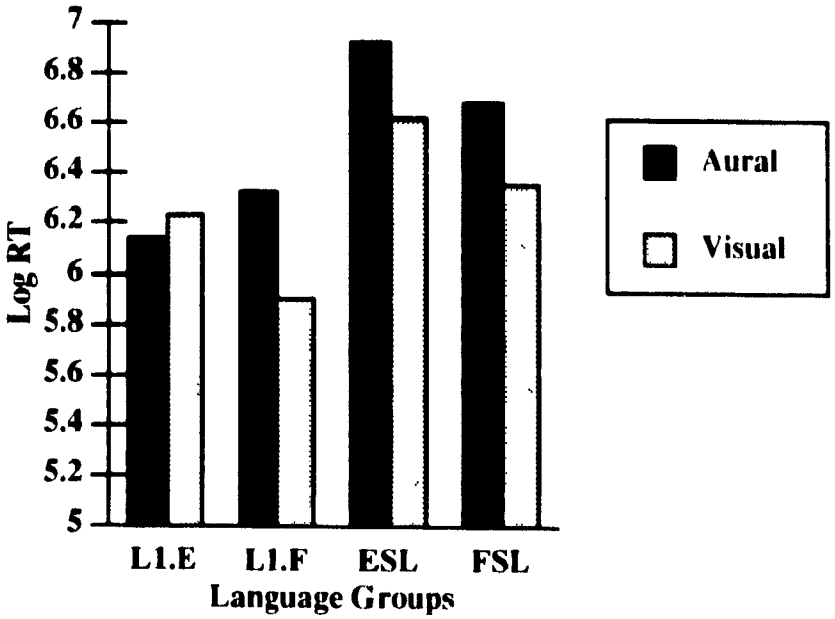


Figure 10 illustrates the effect of modality when the four language groups are compared separately. This interaction is not significant which suggests that the L2 learners are processing the task in a similar way to the native speakers.

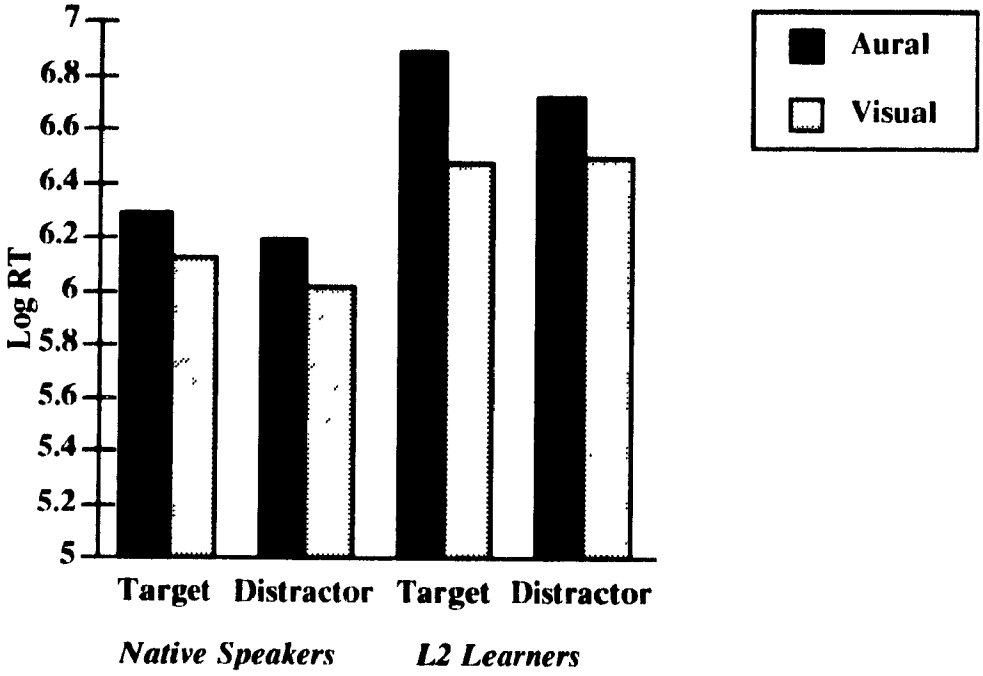
Figure 10. The effect of modality on reaction time to grammaticality judgment for the four language groups.



Certain trends were highlighted in the analysis with respect to modality[8] There was an interaction which approached significance which indicates that L2 learners took longer to judge the sentences aurally than visually [$\underline{F}(1,71) = 3.29$, $p < 0.074$] as shown in Figure 11. Furthermore, the L2 learners took longer overall, than the native speakers [$\underline{F}(1,71) = 17.985$, $p < 0.001$].

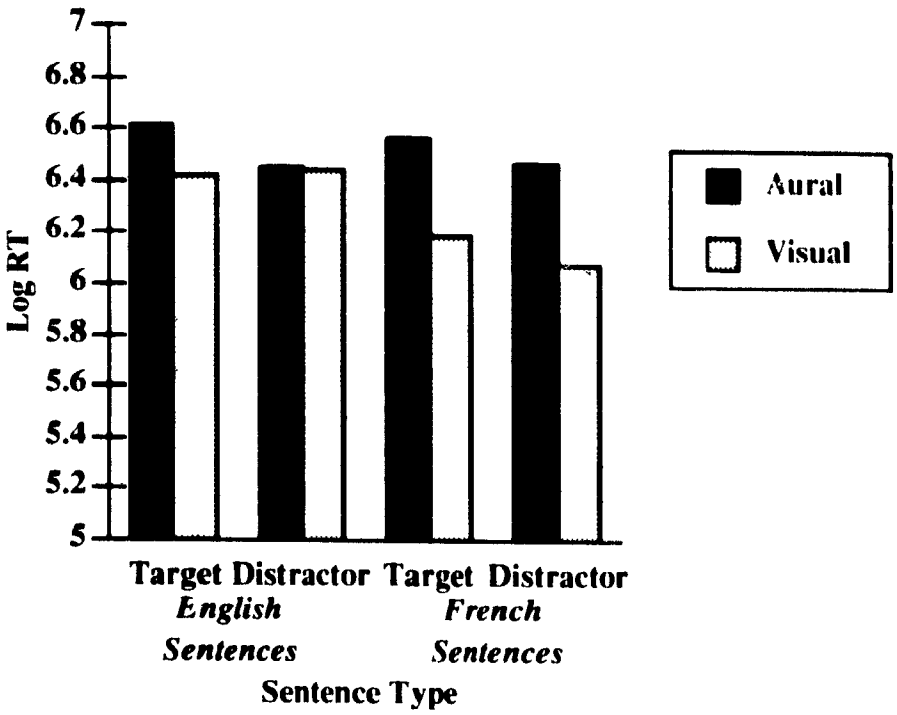
[8]These trends are reported here on the assumption that the relatively small n (10 per cell) is the reason for the failure to reach significance at $\alpha = .05$.

Figure 11. The effect of modality and target for the native speakers and L2 learners on reaction time to grammaticality judgment.



An additional effect which approached significance involves modality with the target effect and the language in which the sentences were presented. Figure 12 illustrates the interaction, where L1.F and FSL subjects combined showed a greater difference in RT between the aural and visual modalities than L1.E and ESL subjects combined [$F(1,71) = 3.71$, $p < 0.058$]. The L1.F and FSL subjects were much faster in the visual modality relative to their performance in the aural modality than were the L1.E and ESL subjects. Furthermore, all subjects had larger RT's for the target sentences in the aural modality than for the targets in the visual modality. Again, the target sentences prompted a larger RT than distractors, and the RT's are longer in the aural than the visual modality.

Figure 12. The effect of modality, target and language of sentences on reaction time to grammaticality judgment.



An inspection of the data which deals with the target sentences (and not modality) reveals that overall, all subjects were slower to judge the ungrammatical subjacency violations than the equivalent grammatical sentences, as illustrated in Figure 13, [$F(1,71) = 12.70$, $p < 0.001$]. While the subjects were slower to judge subjacency violations, this effect did not interact with modality. There is no speed-accuracy trade-off in judging the subjacency violations.

Figure 13. The effect of target and grammaticality on reaction time to grammaticality judgment.

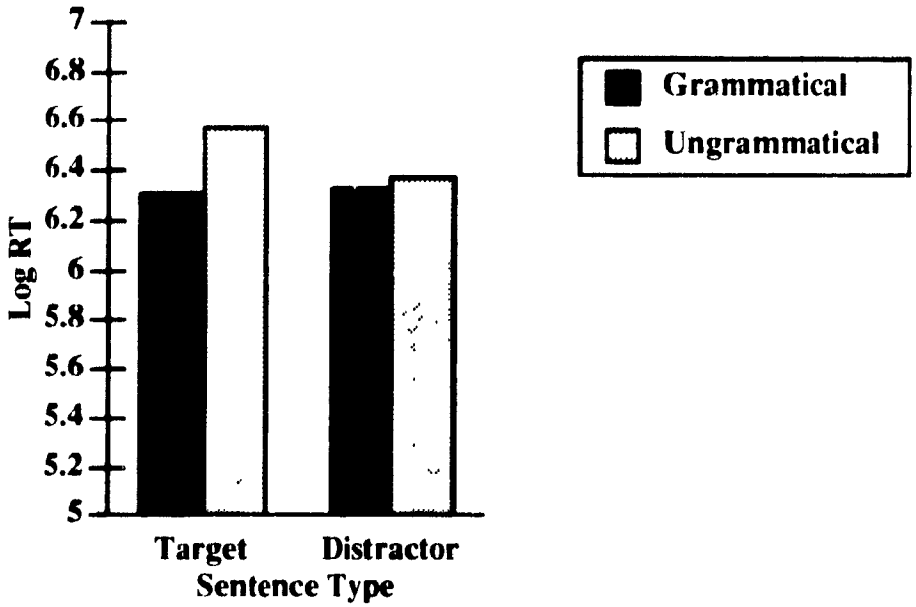
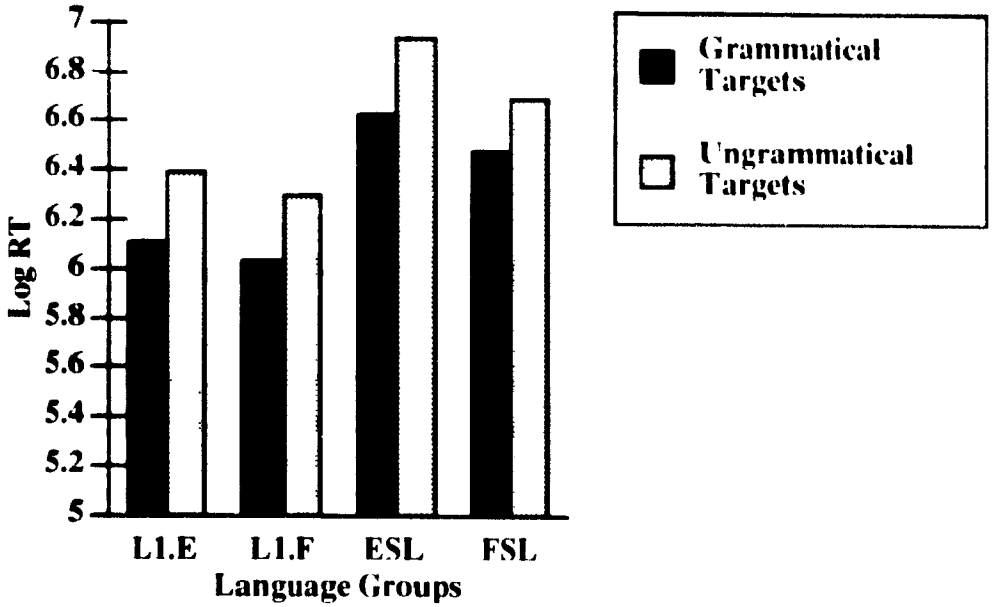


Figure 14 illustrates responding of all four language groups. These interactions are not statistically significant. As in the accuracy analysis, when native speaker and L2 learners data are collapsed, there are significant differences in RT for grammaticality. When the four language groups are compared however, there are no significant differences in RT.

Figure 14. The effect of grammaticality and target for the four language groups.



This non-significant interaction, again implies that whatever processes or knowledge that is being used and accessed to judge these sentences is comparable for subjects across all four language groups. If one assumes that knowledge of UG is mediating the native speakers' judgments, then one could conceivably argue that the L2 learners are also using knowledge of principles of UG to make their judgments. At the very least, however, the L2 learners and native speakers are processing the stimuli in a parallel or similar fashion.

In summary, there is an overall effect of modality in the RT data, since subjects were slower to judge sentences which they heard than those which they read. L2 learners are more affected by the processing demands of the auditory task than the native speakers, especially when judging the subjacency violations. The L1.F and F.SL subjects showed a greater difference between speed of judging the sentences in the aural vs. the visual modality than did the L1.E and ESL subjects. Subjacency violations overall took longer to be judged ungrammatical than either the grammatical sentences with embedded questions or the distractors. As in the accuracy results, when the four language groups

are not collapsed, there is no interaction of language of subjects and grammaticality of the target sentences. Again, this result may suggest that native speakers and L2 learners are processing and accessing similar types of information when making their grammaticality judgments. The type of knowledge accessed by all subjects may include knowledge of principles of UG, suggesting that L2 learners do have access to UG, when performing on such a metalinguistic task.

5.0 Discussion

There were two basic issues addressed in the present experiment. The first issue was the modality of stimulus presentation. Based on the few experiments which had previously examined the issue of modality of stimulus presentation, the hypothesis in the present experiment was that subjects would be slower and less accurate to judge sentences in the aural than the visual modality. This hypothesis was supported. The second issue concerned the subjacency violations. The critical sentences presented to the subjects were those involving wh-islands. Half of these target sentences were grammatical declarative sentences involving embedded questions, while the other half were ungrammatical questions which violated the principle of subjacency. Would subjects respond differently to these target sentences than any others? Furthermore, would subjects respond differently to the ungrammatical subjacency violations than to the grammatical wh-island sentences? The results revealed that subjects were less accurate and slower to judge the subjacency violations than the other grammatical target and distractor sentences. The issue of whether or not Universal Grammar is available to adult L2 learners has not yet been resolved (White, 1989). An accepted belief is that native speakers make their grammaticality judgments on the basis of knowledge of principles of UG. This is one of the reasons why native speakers are often used as a control group in experiments which examine the role of UG in adult SLA. If adult L2 learners are less accurate (than natives) in judging sentences which violate principles of UG, then one can potentially argue that UG is not available to those L2 learners. Correspondingly, if native

speaker and L2 learner performance on the sentences which violate principles of UG is comparable, then one could argue that UG is accessible to the L2 learners. The results of the present experiment revealed that there was no significant interaction between language of subjects (the four groups) and grammaticality of target sentences (the UG violations) on both the accuracy and the reaction time measure[9].

With reference to modality effects, the results clearly indicate that subjects do respond differently in a grammaticality judgment (GJ) task when the sentences are presented in the visual versus the aural modality Modality influences accuracy on ungrammatical sentences since subjects were less accurate at judging ungrammatical sentences when they were heard, than when they were read (see Figure 3). One might have predicted that this effect of modality would only be manifested in the performance of L2 learners since they are presumably less fluent and less familiar with the target language. However, the effect of modality on accuracy on a GJ task was true of both native speakers and L2 learners. Thus, the effect of modality cannot simply be attributed to a function of being an L2 learner. Ungrammatical sentences were judged less accurately than grammatical sentences in the visual condition also, but the effect was much more pronounced in the aural condition. With respect to the RT data, there was clear support for the hypothesis that subjects would be slower in the aural condition than the visual (see Figure 9) and burdens of auditory processing seem to produce a greater obstacle for L2 learners to overcome than for native speakers. As illustrated in Figure 11, this effect is highlighted for the target, wh-island sentences for all language groups, but particularly for the L2 learners.

The present results are compatible with those of Haig (1991) and Johnson (1992) who found that adult learners had significantly lower accuracy scores in the aural condition than the visual one. Furthermore, as in the present experiment, Haig's (1991)

---

[9] See Figures 6 and 14.

L2 learners were particularly less accurate at correctly rejecting violations of subjacency in the aural condition than the visual one. Haig's L2 subjects were native speakers of Chinese while the present experiment tested native speakers of French (on English sentences) and native speakers of English (on French sentences). The two experiments together (Haig's and the present one) offer converging evidence to support the hypothesis that modality will affect subjects performance on a GJ task, with specific effects on subjacency violations.

Thus, not only are there clear modality effects in both the accuracy and RT data, there are also clear effects for the target sentences. The demands of auditory processing affected responding to the target sentences more than the other sentences in the experiment (see Figure 11 and 12). Furthermore, in both the accuracy and the RT analysis, it is clear that subjects as a whole were less accurate and slower at judging subjacency violations than other sentences, in both modalities (see Figure 5 and 13).

It must be noted however, that all groups of subjects were capable of judging the subjacency violations. As indicated in Table 3, the mean score of correct responses on the subjacency violations was 3.75 out of 5 for the ESL subjects and 3.85 out of 5 for the FSL subjects. While these means may not be as high as the means for the native speakers, they are well above a 50% chance accuracy on the crucial target sentences suggesting that learners were not responding randomly.

Given these effects of modality in both the accuracy and speed with which subjects can judge sentences, one can then ask what are the implications of these effects. What do these modality effects mean for researchers investigating adult SLA?

5.1 Implications for research

The modality effects in the GJ tasks are salient on ungrammatical sentences, and specifically those which violate subjacency. Furthermore, the effect of modality most detrimentally affects performance of L2 learners. Researchers investigating L2 grammatical competence using a GJ task are often most crucially interested in L2

learners' judgments of ungrammatical sentences which violate universal principles. The results of the present investigation illustrate how it is precisely these variables which are most affected by the modality of presentation .

When investigating UG in adult SLA, the aim is to attain reliable measures of linguistic competence, to determine what L2 learners 'know' about the target language. Researchers are compelled to make assumptions about linguistic competence from performance data. As Cook (1990, p. 595) stated: "All statements about L2 development depend upon some test of performance; we must always take into account the relevant aspects of performance by native and nonnative speakers before drawing conclusions for the theory of Universal Grammar". Invalid assumptions might occasionally be made concerning learners' underlying competence when relying on performance data. While this problem may not be immediately resolved, it can be constrained. As Birdsong (1989) and Bley-Vroman and Masterson (1989) have suggested, converging and validating evidence is one way of constraining the problems inherent in assuming competence from performance data.

The results of the present study illustrate how converging evidence can provide greater confidence in the conclusions drawn from results attained from empirical research. Overall, subjects in both the aural and visual conditions were less accurate and slower to judge the target sentences. Therefore, one can be confident that even across modalties, subjects were responding differently to the target sentences than the distractors. Furthermore, converging evidence was found across the measures used in this experiment. The results of both the accuracy and the reaction time measures illustrate how modality effects can be found on both types of measure Again, one can be more confident that the conclusion that there is an effect of modality of stimulus presentation is sound since it was manifested on both an accuracy and an RT measure.

With respect to modality therefore, if one can illustrate that processing information in auditory tasks will lead to lower scores on a given measure, then one

might assume that results from visual tasks alone may be an inflated measure of a subjects' grammatical competence. Conversely, results from auditory tasks alone may not accurately reflect a subjects' underlying competence either. As Johnson (1992) has argued, the accounts of the differences in responding to GJ tasks may not be completely obvious. Johnson (1992) suggested that a written task may inflate a subject's performance while an auditory task (and the burdens imposed by on-line processing) might inaccurately depict a lower score on the measure tested.

One way the two presentation modalities may have detrimentally affected performance in both Haig's (1991) and Johnson's (1992) studies is by the differences in presentation time on the written and aural tasks. In both experiments, subjects had more time to process the sentences in the visual condition than the aural condition. The discrepancy in presentation time was particularly salient in Johnson's (1992) study. Any differences in the two modalities may have been a function simply of length of presentation in both Haig's (1991) and Johnson's (1992) research. The present experiment controlled for presentation time. Subjects in the visual condition were restricted in the amount of time in which they could process the sentences, the same amount of time that the subjects in the auditory condition had. Despite the potential for processing burdens in the visual condition (since subjects were limited in the amount of time they had to read each sentence), clear modality effects were still found. The subjects in the visual modality still yielded higher scores of accuracy and faster reaction times. One cannot argue categorically, however, that these differences are due to an inflated metric of grammatical competence in the visual condition and/or a reduced indication of grammatical competence in the auditory task.

The analysis of modality effects in other areas of research, such as short-term memory, indicates that it would be inappropriate to suggest that one modality is a better avenue to underlying grammatical competence than the other. Penney (1989) has convincingly demonstrated that there is a large body of research to show that information

is processed differently depending on the modality within which that information is presented. With reference to GJ tasks, modality effects do not mean that a task in one modality is a better indicator of grammatical competence than the same task in another modality. This could only be possible if linguistic input in one modality yieded superior access to the UG 'module' than the other. Currently, there is no evidence to support such a suggestion. One could hypothesize that if one modality did yield better access to UG-type knowledge, it would be the aural one, since children, as they are learning their L1, are only initially exposed to aural input. However, Haig's (1991), Johnson's (1992) and now the present experiment, all found lower accuracy (and in the present experiment, slower RTs) in the aural modality. A difference in modality does suggest, however, that research which continually presents tasks in one modality, without conducting the same tasks in the other, may not be getting a complete picture of what processes or knowledge subjects may have available in performing on a metalinguistic task.

Why would there be such a difference in responding across the two modalities? Human beings are considered to be limited-capacity processors (Newell & Simon, 1972). While receiving aural linguistic input, listeners have a limited capacity for processing what they hear in the time available (Foss & Lynch, 1969; Foss, 1969; Aaronson, 1974; Green, 1977). If speech input is too fast, or if the listener is not a native speaker of the language or if the input competes with other input, humans' ability to deal with speech breaks down (Clark & Clark, 1977).

Listeners try to isolate and identify surface constituents and hold them in working memory as units as the input is coming in. While trying to construct propositions and representations from the first constituent, the remainder of the sentence also has to be processed. Since all these processes take time, and human beings are limited capacity processors, these processes may interfere with other mental activity and thus cause subjects to be slower and/or less accurate at processing linguistic information aurally (Clark & Clark, 1977). Furthermore, if processing is made difficult by either lexical

entries [by presenting infrequent lexical items (Foss, 1969)] or surface structure [by presented self-embedded sentences (Foss & Lynch, 1969)] reaction times will be slower. Obviously, such working memory limitations are less critical when reading sentences, since readers can look back at surface constituents as they are processing sentences.

Subjacency violations presumably affect the subjects in a similar way as the research described in Foss (1969) and Foss and Lynch (1969), in that processing may be made more difficult since subjects encounter sentences which have a structure which is unpredictable and unfamiliar. These difficulties could undoubtedly cause subjects to judge subjacency violations less accurately and more slowly in an aural task than a visual one. Correspondingly, if subjects **can** successfully perform a task (and overcome processing loads) when stimuli are heard rather than read, then a stronger case for UG's availability can be made. By looking at the means presented in Table 3 and 4, it is obvious that the subjects in the present experiment are able to successfully complete the task in the aural modality.

Comparing tasks presented in different modalities is one means of acquiring validating and converging evidence. As discussed above, if subjects can perform successfully in an auditory task, in addition to a visual one, more positive conclusions can be drawn about whether principles of UG mediate adult L2 acquisition. Analogous to the benefits of using two presentation modalities is the benefit of using two measures of linguistic skill. Different measures can be used to test subjects' knowledge and processing skills of their second language. In the present experiment, both accuracy and RT were measured. In many previous experiments, accuracy alone was the metric from which inferences were drawn concerning subjects' knowledge of their L2 (Haig, 1991; Johnson, 1992; Johnson & Newport, 1991). In line with Birdsong's (1989) suggestions about converging evidence, results from two measures can be more useful.

In the present experiment, modality effects were found both in the accuracy and in the RT data. One can be more confident, therefore, in concluding that modality really

does affect subjects' responses on a GJ task. Having two measures of a subject's performance on a grammaticality task can potentially serve to isolate the kind of knowledge into which the GJ task is tapping. For example, if native speakers and L2 learners are the same on an accuracy measure for a GJ task, but different on an RT measure, then one could argue that the difference between L2 learners and natives may not be one of a paucity in L2 competence [as some may argue, (e.g., Bley-Vroman, 1990)], but rather, be a manifestation of an L2 processing constraint. Without an RT measure in conjunction with an accuracy measure, this possibility could not be investigated. Correspondingly, if native speakers and L2 learners were the same on an RT measure for a GJ task, but different on an accuracy measure, one could either conclude that a) the L2 learners sacrificed accuracy for speed or b) that there was an underlying difference in L2 competence which is not reflected in processing. A further possibility of course, would be if native speakers and L2 learners were both different on an accuracy measure **and** an RT measure. One could conceivably conclude from this type of result that the subjects both a) have different knowledge representations concerning the L2 and b) this difference is manifested in on-line processing tasks. Having only accuracy as a measure on a GJ task would not permit the possibility of contemplating the effect of processing and may cloud any conclusions concerning L2 learners' linguistic and metalinguistic skills. It is highly desirable to have both measures.

The validity of using both accuracy and RT as measures on a GJ task has significant implications for empirical research. White and Genesee (1992), for example, found that their near-native subjects were no different from natives in their responses on both accuracy and RT measures on a GJ task. It was significant for their investigation to have both measures on their GJ task, since data from both measures enabled them to conclude that not only did their subjects have a similar underlying L2 competence but they also processed the L2 in a parallel fashion. Having these two measures of a

subjects' performance on a GJ task supplements their conclusions that UG is available to L2 learners.

An obvious use for RT measures comes in investigating processing strategies (Clahsen, 1984) to account for L1 and L2 differences (Eubank, 1991). Other researchers, use RT as a measure of competence (in conjunction with measures of accuracy) using sentence matching procedures. Bley-Vroman and Masterson (1989) for example, used reaction time in a sentence matching task to investigate L2 learners knowledge of the Functional Categories Parameter. Bley-Vroman and Masterson's Korean subjects were much slower at identifying pairs of sentences as identical or not than native speakers. They argue that while native speakers are more rapid to respond, the pattern of responding among the stimulus items is the same for both native speakers and L2 learners. They conclude that the English setting of the Functional Categories Parameter is being reflected in the RT's of the task and that specifically, the L2 learners' responses reflect an English setting. Reaction time, then, can be a useful measure for Bley-Vroman and Masterson's (1989) task since results dervied from RTs may indicate similar patterns of linguistic processing across language groups[10].

Accuracy and reaction time may not tap into the same underlying system. If a subject is able to correctly judge a sentence that is ungrammatical, one often infers that that subject 'knows' about a given universal principle. What does it mean for a subject to be slow at judging the sentence? If a subject is slower at judging sentences, and yet is still accurate at doing so, what can this inform researchers about an L2 learner's underlying grammatical competence? With specific reference to subjects' ability to judge subjacency violations, their slower RT's may simply be an artifact of the processing limitations discussed above. The difficulty of the surface structures of these sentences makes it difficult for the L2 learners to identify and keep constituents in working

[10]Though there are other problems inherent in a sentence matching task which cannot be discussed in the present context.

memory, thereby resulting in a slower RT. A further possibility is that native speakers might judge the grammaticality of sentences faster than L2 learners because of a familiarity effect since native speakers have had more practice and exposure to the stimuli.

Ideally, researchers would like to refine theories of language acquisition from the results of experiments which examine underlying grammatical competence and not simply processing differences[11]. If on-line processing detrimentally affects L2 learners' performance on any given task, one would not like to draw conclusions about grammatical competence based solely on processing differences. If, however, one can show that processing is the same for a group of native speakers **and** L2 learners, then one has a potentially stronger claim about whether specific principles of UG are instantiated in a subjects' interlanguage grammar.

The present experiment found that when the four language groups were compared, there were no significant interactions with grammaticality on either the accuracy measure or RT (compare Figures 5 and 14). White and Genesee (1992) also found that their L2 subjects had similar accuracy and RT means as their native subjects. One can argue from the RT results that their subjects processed the linguistic information in the same way as natives.

In summary therefore, using different means of testing subjects' knowledge of syntax is a good strategy to attain dependable evidence Using different measures like accuracy and reaction time, and conducting experiments that present tasks in different modalities are some of the ways more reliable evidence might be collected.

5.2 Conclusions

Methodology is a deciding factor in determining the outcomes of research. The present experiment illustrated how the modality within which stimuli are presented in a

[11]Unless, of course, the point of the research is to demonstrate a processing difference

GJ task will influence subjects' accuracy and reaction times at judging sentences which violate principles of UG. Subjects were slower and less accurate to judge violations of subjacency in the aural modality than the visual one. Researchers can take advantage of modality differences, and methodology in general, by conducting research in a variety of contexts (i.e., using different tasks, different principles of UG, etc.). Evidence which finds subjects' sensitivity to principles of UG in both the aural and visual modality is more compelling than just in the visual one. Furthermore, evidence from on-line processing tasks which illustrate sensitivity to principles of UG by L2 learners are more compelling than tasks where subjects can take as much time as they like.

Evidence from a variety tasks, subjects and stimuli, will help researchers determine the precise nature of the role UG plays in adult L2 acquisition. A clearer understanding of the kinds of knowledge to which L2 learner's are sensitive will allow L2 pedagogues to implement the necessary tasks and focus in the language classroom, in turn, enabling L2 learners to successfully acquire their language. A keen understanding of the role methodology plays in L2 acquisition research will assist researchers in obtaining the sound evidence required to determine the information and processes which guide successful L2 acquisition.

References

Aaronson, D. (1974). Stimulus factors and listening strategies in auditory memory: A theoretical analysis. Cognitive Psychology, 6, 108-132.

Adjémian, C. & Liceras, J. (1984). Accounting for adult acquisition of relative clauses Universal Grammar, L1 and structuring the intake. In F Eckman, L. Bell, and D Nelson (Eds.), Universals of second language acquisition. Newbury House Rowley, MA.

Bardovi-Harlig, K. (1986). Markedness and Salience in second-language acquisition Language Learning, 37, 385-407.

Birdsong, D. (1989). Metalinguistic performance and Interlinguistic Competence. New York: Springer.

Birdsong, D. (1992). Knowledge of ungrammaticality in SLA Theory Paper Presented at B.U. Conference on Language Development. October 1992.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. Linguistic Analysis, 20, 3-49.

Bley-Vroman, R., Felix, S., & Ioup, G. (1988). The accessibility of Universal Grammar in adult language learning. Second Language Research, 4, 1-32.

Bley-Vroman, R. and Masterson, D. (1989). Reaction time as a supplement to Grammaticality Judgements in the investigation of second language learners' competence. University of Hawaii Working Papers, 9/2, 207-237.

Broadbent, D., Vines, R., & Broadbent, M. (1978). Recency effects in memory as a function of modality of intervening events. Psychological Research, 40, 5-13.

Chaudron, C. (1983). Research on metalinguistic judgments: A review of theory, methods and results. Language Learning, 33, 343-377.

Chomsky, N. (1965). Aspects of the Theory of Syntax. MIT Press: Cambridge, MA.

Chomsky, N. (1981a). Lectures on Government and Binding. Dordrecht: Foris.

Chomsky, N, (1981b). On the representation of form and function. The Linguistic Review, 1, 3-40.

Clahsen, H. (1984). The acquisition of German word order. In R. Anderson (Ed.), Second Languages. Rowley, MA: Newbury House.

Clahsen, H. & Muysken, P (1986). The availability of universal grammar to adult and child learners: a study of the acquisition of German word order. Second Language Research, 2, 93-119.

Clark, H H. & Clark, E V (1977). Psychology and Language: An Introduction to Psycholinguistics. New York: Harcourt Brace Jovanich.

Coook, V. (1990) Timed comprehension of binding in advanced L2 earners of English. Language Learning, 40:4, 557-599.

Corder, S.P. (1967). The significance of learners errors. IRAL, 5/4, 161-170.

Dennis, I. (1977). Component problems in dichotic listening. Quarterly Journal of Experimental Psychology, 29, 437-450.

Educational Testing Service. (1993). TOEFL (Test of English as a Foreign Language). Princeton: ETS

Eubank, L. (1991). Sentence matching and processing in L2 development. Paper presented at the Conference on Theory Construction, Michigan State University, October 1991.

Finer, D. (1991). Binding parameters in second language acquisition. In L. Eubank, (Ed.), Point Counterpoint. John Benjamins: Amsterdam.

Flynn, S. (1984). A Universal in L2 Acquisition based on a PBD Typology. In F. Eckman, L. Bell & D. Nelson (Eds.), Universals of Second Language Acquisition. Newbury House: Rowley, MA.

Foss, D.J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. Journal of Verbal Learning and Verbal Behaviour, 8, 457-462.

Foss, D.J. & Lynch, R.H. (1969). Decision processes during sentence comprehension Effects of surface structure on decision times Perception and Psychophysics, 5, 145-148.

Frick, R.W. (1984). Using both an auditory and a visual short-term store to increase digit span. Memory & Cognition, 12, 507-514.

Green, D.W. (1977) The immediate processing of sentences Quarterly Journal of Experimental Psychology, 29, 1-12.

Gum, T. & Bub, D. (1988). Psychlab, v.0.85, McGill University, Montreal, Quebec: Montréal Neurological Institute, Neuropsychology

Haegeman, L. (1991). Introduction to Government and Binding Theory. Basil Blackwell Ltd: Cambridge MA.

Haig, J. (1991). Universal grammar and second language acquisition The influence of task type on late learner's access to the subjacency principle TESL Monograph McGill University.

Johnson, J.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, J.S., & Newport, E.L. (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, J.S. & Newport, E.L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language Cognition, 39, 215-258.

Kellerman, E. (1985). Dative alternation and the analysis of data: a reply to Mazurkewich. Language Learning, 35, 91-106.

Kucera, H., & Francis, W.N. (1967). Computational analysis of present-day American English. Providence: Brown University Press.

Lenneberg, E. (1967). Biological foundations of language. New York: Wiley.

Mazurkewich, I (1984). The acquisition of the dative alternation by second language learners and linguistic theory. Language Learning, 34, 91-109.

Newell, A & Simon, H A (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

Penney, C. G. (1974). Order of report and interference effects in four-channel bisensory memory. Canadian Journal of Psychology, 28, 371-382.

Penney, C. G. (1980). Order of report in bisensory verbal short-term memory. Canadian Journal of Psychology, 34, 190-195.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. Memory and Cognition, 17, 398-422.

Pinker, S. (1989). Learnability and Cognition: The acquisition of argument structure. MIT Press: Cambridge.

Pinker, S. (1990). Language Acquisition. In D.N Osherson & H. Lasnik (Eds.), Language: An invitation to Cognitive Science: Volume 1. MIT Press: Cambridge.

Rönnberg, J., Nilsson, L.G. & Ohlsson, K. (1982). Organization by modality, language and category compared. Psychological Research, 44, 369-379.

Ritchie, W.C. (1978) The right roof constraint in an adult acquired language. In W.C. Ritchie (Ed.), Second language acquisition research: issues and implications. New York: Academic Press.

Rizzi, L. (1982). Violations of the wh-island constraint and the subjacency condition. In L. Rizzi Issues in Italian Syntax. Dordrecht: Foris. 49-76.

Rock, I. & Victor, J. (1964). Vision and touch: An experimentally created conflict between the senses. Science, 143, 594-596.

Rollins, H.A., Jr. & Hendricks, R. (1980). Processing of words presented simultaneously to eye and ear. Journal of Experimental Psychology: Human Perception & Performance, 6, 99-109.

Schachter, J. (1989). Testing a proposed universal. In S. Gass & J Schachter (Eds), Linguistic Perspectives on second language acquisition Cambridge Cambridge University Press.

Schmidt, M. (1980). Coordinate structures and language universals in interlanguage Language Learning, 30, 397-416.

Shaffer, L. H. (1975). Multiple attention in continuous verbal tasks In P M.A Rabbitt & S. Dornic (Eds.), Attention and performance V. Academic Press. London

Shallice T., & Warrington, E.K. (1977). Auditory-verbal short-term memory impairment and condition aphasia. Brain & Language, 4, 480-491.

Skinner, B.F. (1957). Verbal Behavior. New York: Appleton-Century Crofts.

Sportiche, D. (1981). Bounding nodes in French. The Linguistic Review, 1, 219-246.

Systat, Inc. (1993). SYSTAT: Statistics, Version 5.2.1. [Computer Program] Evanston, Illinois.

White, L. and Genesee, F. (1992). How native is near-native?: The issue of age and ultimate attainment in the acquisition of a second language. Paper Presented at B.U. Conference on Language Development. October 1992.

White, L. (1985). The acquisition of parameterized grammars: Subjacency in second language acquisition. Second Language Research, 1, 1-17.

White, L. (1986). Implications of parametric variation for adult second language acquisition: an investigation of the 'pro-drop' parameter. In V. Cook (Ed.), Experimental approaches to second language acquisition. pp. 55-72. Pergamon Press: Oxford.

White, L. (1988). Island effects in second language acquisition. In S.Flynn & W. O'Neil (Eds.), Linguistic theory in second language acquisition. Dordrecht: Kluwer.

White, L. (1989). Universal Grammar and Second Language Acquisition John Benjamins: Amsterdam/Philadelphia.

APPENDIX A

Stimuli

English Sentences

Target Sentences

Wh-Ungrammatical Questions

*What did the teacher know why the student said?
*What did the king understand why the enemy revealed?
*What did the professor guess why the writer wrote?
*What did the family forget why the boy kicked?
*What did the university understand why the team drank?
*What did the lady know why the man bought?
*What did the woman wonder why the baby dropped?
*What did the parents discover why the children played?
*What did the attorney realize why the people robbed?
*What did the patient know why the doctor sent?

WH-Grammatical declaratives

The teacher knew why the young student said hello.
The king understood why the enemy revealed the plan.
The professor guessed why the writer wrote the book.
The family forgot why the boy kicked the girl.
The university understood why the team drank the champagne.
The lady knew why the man bought the car.
The woman wondered why the baby dropped the toy.
The parents discovered why the children played the music.
The attorney realized why the bad people robbed banks.
The patient knew why the doctor sent the medicine.

Distractor Trials

Distractor - Declarative Ungrammatical
*The schools do admire the principal who history teaches.
*The judge went to the store the surprised manager.
*The detective who finds hate the congress many mistakes.

Distractor - Question Grammatical
What did the parents think the young children wanted?
What did the doctor believe the new nurse said?
What did the police officer think the thieves took?

Distractor - Declarative Grammatical
The couple spoke with the person who lives downstairs.
The chairman doesn't like the director who was rude.

Distractor - Question Ungrammatical
*When the analyst did remember that the appointment was.

*Who the secretary did know that the boss fired?

First 5 and last 3 dummy experimental trials

The pupil felt upset at being pressured in class.
What did the nurse think about the new patient?
*When did the old carpenter wish cabinet finished soon?
*All night long Jane worried and the called hospital.
*The dinner party was very new her exciting for .

The lawyer drinks too much coffee in the morning.
*What didn't the child like the meal offered at?
Does the banker laugh at the poor loan applicants?

12 practice trials

John believed that he would soon pass the test.
Jane went for a walk in the large field.
Jane's mother told her she could never see John.
*At her ate party Jane two pieces of cake.
*John read the book was detailed very and long .
*When important it was the computer always broke down.
Why didn't he understand the story told to him?
Was the man at the window someone you knew?
Couldn't the husband see that his wife was upset?
*Why did the dress Jane feel compelled to wear?
*How did she manage to herself hurt playing there?
*What did the farmer red park the tractor in?

French Sentences

Target Sentences

Wh-Ungrammatical Questions

*Qu'est-ce que le professeur a su pourquoi l'étudiant a dit?
*Qu'est-ce que le roi a compris pourquoi l'ennemi a révélé?
*Qu'est-ce que le professeur a deviné pourquoi l'écrivain a écrit?
*Qu'est-ce que la famille a oubliée pourquoi le garçon a frappé?
*Qu'est-ce que l'université a compris pourquoi l'équipe a bu?
*Qu'est-ce que la dame a su pourquoi l'homme a acheté?
*Qu'est-ce que la femme s'est demandé pourquoi le bébé a échappé?
*Qu'est-ce que les parents ont découvert pourquoi l'enfant a joué?
*Qu'est-ce que l'avocat a compris pourquoi les gens ont volés?
*Qu'est-ce que le patient a su pourquoi le docteur a envoyé?

WH-Grammatical declaratives

Le professeur savait pourquoi le jeune étudiant a dit allo.
Le roi a compris pourquoi l'ennemi a révélé le plan.
Le professeur a deviné pourquoi l'écrivain a écrit le livre.
La famille a oublié pourquoi le garçon a frappé le fille.
L'université a compris pourquoi l'équipe a bu du champagne.
La dame savait pourquoi l'homme a acheté l'automobile.

La femme se demandait pourquoi le bébé a échappé le jouet.
Les parents ont découvert pourquoi l'enfant a joué de la musique.
L'avocat a compris pourquoi les bandits volent les banques.
Le patient savait pourquoi le docteur a envoyé le médicament.

Distractor Trials

Distractor - Declarative Ungrammatical
*Les écoles admirent le principal qui histoire enseigne.
*Le juge a été au magasin le a surpris directeur.
*Le détéctive qui trouve hait le parlement plusieurs erreurs.

Distractor - Question Grammatical
Qu'est-ce que les parents pensent que le jeune enfant voulait?
Qu'est-ce que le docteur croit que la nouvelle infirmière a dit?
Qu'est-ce que le policier pense que le voleur a pris?

Distractor - Declarative Grammatical
Le couple a parlé avec la personne qui vit au premier étage.
Le grand patron n'aime pas le directeur qui a été dur.

Distractor - Question Ungrammatical
*Quand l'analyste est-ce que s'est rappelé que le rendez-vous était?
*Qui la sécretaire est-ce que a su que le patron a renvoyé?

First 5 and last 3 dummy experimental trials
L'élève s'est senti affecté par la pression en classe.
Qu'est-ce que l'infirmière pense du nouveau patient?
*Quand est-ce que le charpentier espérait cabinet a fini bientôt?
*Toute la nuit durant Jeanne était préoccupé et la a appellé hôpital.
*La réception était très nouvelle elle excitante pour.

L'avocat boit trop de café le matin.
*Qu'est-ce que l'enfant n'a pas aimé le repas offert à?
Est-ce que le banquier rit des pauvres demandeurs de prêts?

12 practice trials
Jean croyait qu'il pourrait passer l'examen bientôt.
Jeanne a pris une marche dans le grand champ.
La mère de Jeanne lui a dit qu'elle ne pourra jamais voir Jean.
*A son a mangé réception Jeanne deux morceaux de gâteau.
*Jean a lu le livre était détaillé très et long.
*Quand important il était l'ordinateur toujours brisé.
Pourquoi est-ce qu'il n'a pas compris l'histoire qui lui a été raconté?
Est-ce que l'homme devant la fenêtre est quelqu'un que tu a connu?
Le mari ne pouvait-il pas voir que sa femme était boulversée?
*Pourquoi est-ce que la robe Jeanne se sentait poussé à porter?
*Comment est-ce qu'elle a fait se faire mal en jouant là-bas?
*Qu'est-ce que le fermier rouge stationné le tracteur dans?

APPENDIX B

Summary Source Table for Accuracy Analysis A.

UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS

BETWEEN SUBJECTS

SOURCE	SS	DF	MS	F	P
MOD	0.253	1	0.253	0.240	0.626
LANGUE	0.153	1	0.153	0.145	0.705
NATIVE	6.328	1	6.328	5.989	0.017
MOD*LANGUE *NATIVE	0.253	1	0.253	0.240	0.626
MOD*LANGUE	3.003	1	3.003	2.842	0.096
MOD*NATIVE LANGUE	3.003	1	3.003	2.842	0.096
*NATIVE	0.028	1	0.028	0.027	0.871
ERROR	76.075	72	1.057		

WITHIN SUBJECTS

SOURCE	SS	DF	MS	F	P
target	1.653	1	1.653	2.502	0.118
target*MOD target	0.903	1	0.903	1.367	0.246
*LANGUE target	0.028	1	0.028	0.043	0.837
*NATIVE target*MOD *LANGUE	2.628	1	2.628	3.977	0.050
*NATIVE target*MOD	0.078	1	0.078	0.118	0.732
*LANGUE target*MOD	1.128	1	1.128	1.707	0.195
*NATIVE target *LANGUE	0.003	1	0.003	0.005	0.945
*NATIVE	0.253	1	0.253	0.383	0.538
ERROR	47.575	72	0.661		

gramm	7.503	1	7.503	11.379	0.001
gramm*MOD	3.828	1	3.828	5.806	0.019
gramm*LANGUE	3.003	1	3.003	4.555	0.036
gramm*NATIVE	5.778	1	5.778	8.763	0.004
gramm*MOD *LANGUE *NATIVE	0.078	1	0.078	0.118	0.732
gramm*MOD *LANGUE	0.528	1	0.528	0.801	0.374
gramm*MOD *NATIVE	0.028	1	0.028	0.043	0.837
gramm*LANGUE *NATIVE	0.028	1	0.028	0.043	0.837
ERROR	47.475	72	0.659		
target*gramm	4.278	1	4.278	10.074	0.002
target*gramm *MOD	0.253	1	0.253	0.596	0.443
target*gramm *LANGUE	0.003	1	0.003	0.007	0.932
target*gramm *NATIVE	0.253	1	0.253	0.596	0.443
target*gramm *MOD*LANGUE *NATIVE	0.528	1	0.528	1.244	0.268
target*gramm *MOD*LANGUE	1.378	1	1.378	3.245	0.076
target*gramm *MOD*NATIVE	0.903	1	0.903	2.127	0.149
target*gramm *LANGUE *NATIVE	0.078	1	0.078	0.184	0.669
ERROR	30.575	72	0.425		

APPENDIX C

Summary Source Table for Reaction Time Analysis A.

UNIVARIATE AND MULTIVARIATE REPEATED MEASURES ANALYSIS

BETWEEN SUBJECTS

SOURCE	SS	DF	MS	F	P
MOD	4.576	1	4.576	4.324	0.041
LANGUE	2.037	1	2.037	1.925	0.170
NATIVE	19.031	1	19.031	17.985	0.000
MOD*LANGUE *NATIVE	1.294	1	1.294	1.223	0.272
MOD*LANGUE	1.507	1	1.507	1.424	0.237
MOD*NATIVE LANGUE *NATIVE	0.399	1	0.399	0.377	0.541
ERROR	0.592	1	0.592	0.559	0.457
	75.127	71	1.058		

WITHIN SUBJECTS

SOURCE	SS	DF	MS	F	P
target	0.621	1	0.621	11.521	0.001
target*MOD	0.154	1	0.154	2.862	0.095
target *LANGUE	0.028	1	0.028	0.515	0.476
target *NATIVE	0.021	1	0.021	0.384	0.538
target*MOD *LANGUE *NATIVE	0.051	1	0.051	0.943	0.335
target*MOD *LANGUE	0.200	1	0.200	3.710	0.058
target*MOD *NATIVE	0.177	1	0.177	3.287	0.074
target *LANGUE *NATIVE	0.075	1	0.075	1.393	0 242
ERROR	3.828	71	0.054		

gramm	1.913	1	1.913	12.680	0.001
gramm*MOD	0.043	1	0.043	0.287	0.594
gramm*LANGUE	0.039	1	0.039	0.256	0.614
gramm*NATIVE	0 001	1	0.001	0.006	0.939
gramm*MOD *LANGUE *NATIVE	0.078	1	0.078	0.519	0.474
gramm*MOD *LANGUE	0.007	1	0.007	0.048	0.827
gramm*MOD *NATIVE	0.012	1	0.012	0.083	0.775
gramm*LANGUE *NATIVE	0.007	1	0.007	0.047	0.830
ERROR	10.709	71	0.151		
target*gramm	0.952	1	0.952	12.700	0.001
target*gramm *MOD	0.074	1	0.074	0.988	0.324
target*gramm *LANGUE	0.003	1	0.003	0.044	0.834
target*gramm *NATIVE	0.000	1	0.000	0.003	0.958
target*gramm *MOD*LANGUE *NATIVE	0.144	1	0.144	1.918	0.170
target*gramm *MOD*LANGUE	0.005	1	0.005	0.069	0.793
target*gramm *MOD*NATIVE	0.012	1	0.012	0.166	0.685
target*gramm *LANGUE *NATIVE	0.060	1	0.060	0.803	0.373
ERROR	5.321	71	0.075		