THE LEXICO-GRAMMATICAL COMPLEXIFICATION LEVEL OF THE INTERLANGUAGE OF BRAZILIAN BEGINNING EFL STUDENTS: AN EXPLORATORY STUDY USING SYSTEMIC-FUNCTIONAL GRAMMAR AS A DATA CATEGORIZATION MODEL[1]

PEDRO HENRIQUE LIMA PRAXEDES FILHO[2]
(Universidade Estadual do Ceará – UECE)

RESUMO

O objetivo deste estudo é investigar, sob a perspectiva da hipótese da IL inicial (Ellis, 1982), o nível de complexificação léxico-gramatical da IL português-inglês de aprendizes principiantes. Quatro rapazes e seis moças, estudantes universitários e matriculados no Semestre 2 da Casa de Cultura Britânica/UFC, foram os sujeitos que produziram, de improviso, narrativas orais e escritas sobre uma experiência pessoal marcante. Com base na GSF, as narrativas foram segmentadas em orações hierarquizadas e sub-hierarquizadas, as quais foram categorizadas quanto às funções configuracionais realizadoras dos sistemas de transitividade e modo. Considerando-se o critério de que um nível baixo de complexificação significa a ocorrência de até 50% de orações hierarquizadas completas quanto à presença simultânea de todas as funções configuracionais realizadoras dos dois sistemas, a hipótese de que as narrativas orais e escritas apresentam um nível baixo de complexificação foi confirmada para as narrativas orais separadamente e as narrativas orais e escritas conjuntamente, tendo sido refutada para as narrativas escritas separadamente. A hipótese de que as narrativas escritas, em relação às orais, apresentam um nível de complexificação mais elevado foi confirmada. Uma vez que o projeto maior é uma pesquisa longitudinal objetivando identificar os traços léxico-gramaticais do *continuum* simplificação-complexificação característico do desenvolvimento da IL de aprendizes de L2 no ambiente de sala de aula, este estudo, que demonstrou a adequação da GSF como ferramenta evidenciadora desse *continuum* também para principiantes, recomenda que um outro piloto transversal seja conduzido a fim de que os três níveis de proficiência de uma mesma instituição possam ser estudados simultaneamente.

INTRODUCTION

This paper reports on an investigation whose theme is inserted within the scope of interlanguage (IL) studies. The study described here investigated more specifically the level of lexico-grammatical complexification of spoken and written narratives rendered by beginning EFL learners[3]. The learners who produced the data were chosen among the Brazilian students who were taking the 'Semester 2' stage (S2) of Federal University of Ceará's *Casa de Cultura Britânica* (*CCB/UFC*) in January/2002. The complexification level of the narratives' lexico-

---

[2] Doctoral student in English Language and Applied Linguistcs at *Universidade Federal de Santa Catarina-UFSC*.

[3] The same feature was investigated as to the advanced and intermediate proficiency levels in Study 1 and Study 2 (cf. V 40), respectively.

grammar was quantified after their having been categorized by means of the Systemic-Functional Grammar (SFG) theoretical model.

The research was oriented by the following objectives: 1) to identify the lexico-grammatical complexification level (low, moderate, or high) of the subjects' spoken and written narratives; 2) to find out whether there is any difference as for the level of lexico-grammatical complexification between the subjects' spoken and written narratives.

The categorized narratives were analyzed in view of testing the following working hypotheses: 1) Since the subjects are beginning EFL learners, their spoken and written narratives (both separately and together) bear a low level of lexico-grammatical complexification; 2) Since the writer has more time than the speaker to elaborate on his/her discoursal production, the subjects' written narratives will bear a higher level of lexico-grammatical complexification than their spoken ones.

The assumption that learners' IL develops mainly by means of strategies of simplification and complexification of the L2 lexico-grammar comes from Ellis's (1982) 'initial IL hypothesis'. Their definitions, however, must be SFG-based as this is the chosen analytical tool through which the referred to strategies are made explicit. The definitions that will be utilized in the present study for the purpose of reaching the objectives and testing the hypotheses were constructed empirically in Study 1 and Study 2 (cf. V 40). Figure 1 shows the evolution of the definitions:

STUDY 1 (INITIAL DEFINITION) ↓	STUDY 1 (FIRST ENLARGEMENT OF THE DEFINITION, PROPOSED IN THE 'DISCUSSION') ↓	STUDY 2 (SECOND ENLARGEMENT OF THE DEFINITION, PROPOSED IN THE 'METHODOLOGY') ↓	STUDY 2 (THIRD ENLARGEMENT OF THE DEFINITION, PROPOSED IN THE 'DISCUSSION') ↓
LEXICO-GRAMMATICAL COMPLEXIFICA TION ↓	A TYPOLOGICAL CLASSIFICATION OF SIMPLIFICATION/INCO MPLETENESS ↓	INCORPORATION OF THE DOWN-RANKED INCOMPLETE CLAUSES) ↓	INCORPORATION OF SIMPLIFIED DEVIATIONS ↓
	1- LAP [language acquisition processes] SIMPLIFICATION/INCO MPLETENESS ↓		(_Categories_: Deviant Complete Ranking Clauses = DCC & Simplified Deviant Complete Ranking Clauses = SDCC). Both are complete clauses (all the transitivity and mood configurational functions are present) that contain deviations in relation to the English language norm used in the adopted textbook – educated British English. Whereas in the latter case the deviations are characterized by an attempt, on the part of the subjects, to opt for simpler lexico-grammatical forms in order to instantiate the transitivity and mood configurational functions, in the former case the deviations do not hold this simplifying feature. Only the SDCC category might be of interest to the IL lexico-grammatical simplification/complexification issue. Its incorporation required a
The production, on the part of the subjects, of **ranking clauses** in their spoken and written narratives, **which have all the lexico-grammatical structural slots filled in by the configurational functions that realize the systems of transitivity and mood** as proposed by Halliday (1994). (_Category_: **Complete Ranking Clauses = CC**).	Simplification is a result of the fact that "[...] learners have not yet acquired the necessary linguistic forms [...]" (Ellis, 1994: 89). **Simplified/incomplete ranking clauses are those produced without resorting to self-corrections and hesitations. They are the ACTUAL SIMPLIFICATIONS or INCOMPLETENESSES.** (_Category_: **LAP Incomplete Ranking Clauses = LAP**).	Up to this point, the down-ranked clauses had been left out of the categorization because they are either whole constituents or part of constituents of the ranking clause in which they are embedded. It was this definitional characteristic that led me to realize that **the lexico-grammatical incompleteness of a down-ranked clause makes the ranking clause of which it is a constituent also incomplete, regardless of the latter being complete.** (_Categories_: **LAP Incomplete Ranking Clauses with Down-Ranked-Clause-Related Incompleteness & LPP Incomplete Ranking Clause with Down-Ranked-Clause-Related Incompleteness)**	
LEXICO-GRAMMATICAL	2- LPP [language production processes] SIMPLIFICATION/INCO MPLETENESS ↓		

120

SIMPLIFICATIO N/ INCOMPLETENE SS ↓	Simplification is a result of naturally occurring pauses in spoken language for discourse planning as a consequence of the necessity for correction (message adjustment) or as a consequence of hesitation (Praxedes Filho, 1996: 149-150), or as a consequence of the inability "... to access them [the necessary linguistic forms] in the production of specific utterances" (Ellis, 1994: 89).	Related Incompleteness).	issue. Its incorporation required a broadening of the definition of lexico-grammatical simplification: the inclusion of deviant complete ranking clauses whose deviations can be classified as simplifications in terms of either the transitivity or the mood system. (*Derived Category*: SDCC with Down-Ranked-Clause-Related Incompleteness).
The occurrence of ranking clauses whose configurational functions, for the same two systems, are not all present simultaneously. (*Category*: Incomplete Ranking Clauses = IC).	**Simplified/Incomplete ranking clauses are those produced as a result of self-corrections and hesitations. Since these clauses are idiosyncratically characteristic of spoken discourse for both native and non-native speakers, they are considered as PSEUDO-SIMPLIFICATIONS or PSEUDO-INCOMPLETENESSES.** (*Category*: **LPP Incomplete Ranking Clauses = LPP**).		

Figure 1: The evolution of the definitions of lexico-grammatical simplification/complexification

The main SFG concepts[4] necessary for the understanding of these lexico-grammatical simplification/complexification definitions are summarized in Figure 2:

THE METAFUNCTIONS OF THE SEMANTIC STRATUM		
IDEATIONAL	**INTERPERSONAL**	**TEXTUAL[5]**
EXPERIENTIAL OR REFLEXIVE: The human being is capable of representing the experiences of his/her outer and inner worlds. Content function. Addressor is observer.	**ACTIVE:** The human being acts, within the social reality, upon other people to make them act as well. Interaction function. Addressor is intruder / participant.	**INSTRUMENTAL:** The human being is able to compose cohesive and coherent spoken and written texts.
ARE REALIZED – AT CLAUSE RANK – BY THE SYSTEMS OF THE LEXICO-GRAMMATICAL STRATUM		

[4] For a more detailed review, the reader should refer to Study 2 in V 40.
[5] All that is related to the textual metafunction is outside the scope of the present study.

TRANSITIVITY (clause as representation)		MOOD (clause as exchange)		THEME (clause as message)	
WHICH, IN TURN, ARE REALIZED BY THE FOLLOWING STRUCTURES OF CONFIGURATIONAL FUNCTIONS					
TRANSITIVITY					
Participant		Process		(Participant)	(Circumstance)
I (Carrier)		was (Relational)		in in home (Attribute)	one day at morning[6] (location/time)
MOOD					
Subject	Finite		Predicator	Complement	Adjunct
I	'past'		'be'	in in home	one day at morning
Mood			Residue		
THEME					
Theme			Rheme		
I			was in in home one day at morning		
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX					
RANKING CLAUSES relate themselves with only same-rank units – other clauses Notation: **VERTICAL STROKES →** ‖ ‖ (Halliday, 1989: 66)					
EXAMPLE (invented): ‖ At home, I realized ‖ that my dog had died. ‖ → Two ranking clauses.					
DOWN-RANKED CLAUSES function as constituents or parts of constituents within the structure of the one-rank-below unit – the group/phrase Notation: **DOUBLE SQUARE BRACKETS →** [[ ]] (Halliday, 1989: 71)					
EXAMPLE: ‖ At home I saw [[my dog die.]] ‖[7] → One ranking clause with a down-ranked clause embedded within it.					

Figure 2: A summary of the main SFG concepts

The relevance of the investigatory endeavor reported on here lies in the fact that it fulfills the recommendations made in Studies 1 and 2 as to the need to do a cross-sectional pilot study with beginning EFL learners, so that a more encompassing and effective evaluation of the adequateness of SFG as a data categorization framework in longitudinal investigations aimed at studying the simplification-towards-complexification process of IL could be accomplished. This need is justified by the fact that the aforementioned studies investigated the IL of advanced and intermediate learners, respectively, and there are no studies to date that have investigated the suitability of SFG to categorize IL data produced by beginning EFL learners.

This pilot-study, the third and last in the series, is relevant, after all, as it represents the final step towards the confirmation of the appropriateness of the analytical dimension of the methodological aspect of my Doctoral Research Project, entitled: 'The development of simplified EFL interlanguage discourses towards complexification: a study using systemic functionalism as a data categorization framework'. In my Doctoral Research Project, I propose a longitudinal type of IL investigation. I conclude Study 2 pointing out the importance of conducting a third cross-sectional pilot-study with beginning EFL learners, aiming at completing the testing cycle of the appropriateness of SFG as a data categorization framework for longitudinal IL simplification-complexification studies.

---

[6] Ranking Clause 2 in Spoken Narrative 08 from the data.
[7] Ranking Clause 7 in Written Narrative 10 from the data.

Since the subjects in this last piece of research that completes the trilogy are beginning EFL learners, I see relevance in reviewing some of the most important initial IL hypotheses found in the literature.

1. INITIAL IL HYPOTHESES

In order that the reader can more easily comprehend the initial IL hypotheses reviewed in this section, it is mandatory that I, at first, make explicit what my understanding is, based on the relevant literature, of the definitions for the terms 'initial IL' and 'initial IL hypothesis'. For the sake of further clarification, it is important to state that whereas Ellis (1982: 207) adopts the term 'learner's initial hypothesis', Selinker (1992: 164) makes use of the term 'initial IL hypothesis'. My choice is for the latter since it is, in my view, more transparent as to its signification: 'learner' presupposes a spectrum of different concepts; 'initial IL', on the other hand, presupposes, more straightforwardly, the concepts of 'IL', 'IL Theory'[8] and 'L2 learner'.

'Initial IL' is the very first "mental grammar" (Ellis, 1997: 33) or "stable plateau" (Selinker, 1992: 226) to emerge, for a given learner, within the IL continuum (the learner's "built-in syllabus", in Corder's [1981a: 9] terms) that is formed between his/her L1 and the L2 he/she is trying to acquire[9]. It is, thus, the simplest, the most primitive IL, the one that is farthest from the L2.

Based on the definition for 'initial IL', it is not difficult to infer what the second term, 'initial IL hypothesis', means. It is the precarious and *a priori* answer to the following research question: How does the 'initial IL' emerge? How does it come into being? What is its genesis? How is the acquisition process of a certain L2 initiated for any learner?

Many Second-Language-Acquisition (SLA) researchers have attempted to come up with an answer to this question. Here, however, I will review the proposals put forward by representatives of an exclusively psycholinguistic position – Selinker (1974, 1992), Corder (1978, 1981a, 1981b, 1992), and Nemser (1974) – and by representatives of a mixed socio/psycholinguistic position – Ellis (1982), Schumann (1978, *apud* Ellis, 1985; McLaughlin, 1987), and Andersen (1983, *apud* Ellis, 1985; McLaughlin, 1987).

Before initiating the review, it is necessary to point out that Corder (1978, 1981a, 1981b, 1992) classifies the 'initial IL hypotheses' into two types: restructuring and recreation hypotheses. For the former, the starting point is the full L1 system that is then slowly and successively restructured into the L2. In this case, there is simplification followed by complexification of the L2 system. For the latter, the starting point is a stripped-to-its-bare-essentials version of the L1 system that is then built upon, in a recreation process towards the L2. In this case, there is no simplification of the L2 system;

---

[8] For the definitions of 'IL' and 'IL Theory', the reader should refer to Study 2 (cf. V 40).

[9] Following Ellis (1994), 'to acquire'/'to learn' and 'acquisition'/'learning' are being used interchangeably in this paper. In the Preface to his book, McLaughlin (1985: x) states that "[...] the distinction between *acquisition*, used to refer to naturalistic acquisition of a language, and *learning*, used to refer to learning through formal classroom instruction (Krashen, 1981b) will not be used [...] the terms "acquisition" and "learning" are used interchangeably in this book and carry no theoretical connotations".

there can only be complexification or elaboration of the initial simple version of the L1 system. Corder claims that the beginning learner cannot simplify what he/she does not know yet: the L2. However, Ellis (1994: 89) refutes Corder's claim, postulating that the learner initially simplifies the L2 by "[...] attending to and internalizing [...]" the salient, easy-to-perceive features of the input (a beginning L2 learner's simplified utterance like "Hitting" can be the realization of "He is hitting me" [Ellis, 1994: 90], as a result of his/her having attended to and internalized only the lexical verb due to its greater salience in the input). According to such a classification, Corder's own 'initial IL hypothesis' is of the recreation type and Selinker's, Nemser's, Ellis's, Schumann's, and Andersen's proposals are of the restructuring type.

Inasmuch as Selinker's (1974, 1992) proposal is concerned, the 'initial IL' results from interlingual identifications that start being made when the learner has his/her first contact with L2 input and are processed by a latent psychological structure located within the brain. Based on his/her own perception of the typological distance between his/her L1 and the L2 he/she is attempting to acquire, the learner identifies, in the L2, features that, for him/her, are similar – at the semantic, morphosyntactic, lexical, and phonological levels – to L1 features and equate them as same, which leads to the emergence of the 'initial IL', the first "mental grammar", the first "stable plateau" – a result of simplification strategies applied on the L2[10]. It is, ultimately, a proposal that takes into account solely psycholinguistic processes, i.e., the learner's cognitive apparatus is capable of initiating, all by itself, the L2 acquisition. However, such a powerful cognitive apparatus cannot prevent the fossilization, sooner or later, of the "built-in syllabus", whose following IL stages are, to some extent[11], the result of complexification strategies applied on the initially simplified L2.

Still within the domain of the exclusiveness of the psycholinguistic processes, we come across Corder (1978, 1981a, 1981b, 1992), who claims that, as stated earlier, the source of the 'initial IL' is not the whole of the L1. In other words, the origin of the 'initial IL' is not the thorough knowledge the learner holds of his/her L1. Therefore, Corder (1978, 1981a, 1981b, 1992) postulates that the 'initial IL' source is a possibly universal simple linguistic code – something like the L1's 'skeleton', whose activation is likely related to the remembrances the L2 learner has of his/her experience in acquiring it. Corder (1992: 25) justifies himself saying that "... we all know a simple basic code because we ourselves have created one in the course of acquiring a first language".

---

[10] Hawkins (1998: 112) interprets Selinker's simplification as an L2 learning strategy, defining it as "... the tendency to reduce the target language [L2] to a small set of general properties".

[11] For 95% of the L2 learners, the complexification strategies stop being applied well before the "built-in syllabus" completely merges into the L2 (Selinker, 1974: 34). On the other hand, Johnson (1992: 185/180) claims that her "tennis clinic strategy" for teaching is capable of defossilizing the accumulated linguistic competence of a "fluent-but-fossilized intermediate student". Despite Nickel's (1998: 6) corroboration of Selinker's proposed fossilization figure, he defends the idea that "... there is (...) the possibility for very good teachers to postpone or bypass some effects of fossilization". Odlin (1993: 382) also goes along with Selinker's (1992) claims on fossilization but suggests that, if SLA is to have any relevance for L2 pedagogy, it must "... deal (...) also with the path of progress that highly successful learners have followed". Although Washburn (1994) makes no claim against fossilization as an SLA fact, he states that it still needs to be better understood both theoretically and empirically, especially due to the lack of linguistic criteria for the identification of L2 fossilized speakers.

124

After the return to a sort of primitive L1, the simple basic code – the first IL, the first "mental grammar", the first "stable plateau" – results from the elaboration the L2 learner submits it to. Likewise, the subsequent ILs, components of the continuum towards the L2, result from the permanent elaboration/complexification of the initial basic source[12]. Corder, however, also agrees with the occurrence of fossilization, which means that the complexification process halts, for the great majority of learners, well before the final goal: a native-like L2.

Nemser's (1971, *apud* Selinker, 1992: 176; 1974) proposal is very similar to that suggested by Selinker (1974, 1992) for the 'initial IL'. Upon having the first contact with L2 input, the learner's first "approximative system" (independent of both the L1 and the L2) is created as a consequence of the psycholinguistic process of intersystemic identifications (Nemser's term for Selinker's interlingual identifications). The first "approximative system" is "[...] characterized by the extensive *underdifferentiation* (syncretism)[13] of L_t [L2] phonological, grammatical and lexical categories [...]" (Nemser, 1974: 59) (emphasis is the author's). Along the developmental continuum, the following "approximative systems" ($L_{a...n}$ = "approximative system" 2, 3, 4, etc, equivalent to "... successive stages of proficiency" [Nemser, 1974: 56]) keep growing in complexity as they draw on **three**, instead of on just two, linguistic systems, namely the L1, the L2, and the accumulated L_a ('approximative system') to date. Again, for most learners, especially adults, the developmental continuum is interrupted before the attainment of a native-like L2 due to the action of fossilization[14].

Ellis's (1982) 'initial IL' proposal is focused on the classroom, rather than on naturalistic settings. Besides, he characterizes it as having a "... socio-cognitive ..." nature (1982: 219) and develops it by means of four hypotheses. Despite the fact that Ellis himself does not classify his hypotheses in any manner, I claim that while Hypotheses 1 and 4 bear a social nature, Hypotheses 2 and 3 have a cognitive nature.

Hypothesis 1 was thus formulated:

The L2 learner utilizes his knowledge of the conceptual organization of events and simplifies their representation in the L2 according to principles of informativeness. He operates a strategy of **semantic simplification** (Ellis, 1982: 214-215) (emphasis is mine).

Ellis arrived at this hypothesis by way of observing the communicative need of a beginning L2 learner who does not posses yet the sufficient and necessary L2 linguistic resources in order for him/her to carry out the intended interaction. The learner, hence, "... relies extensively on the situational context 'speaking' for him ..." (Ellis, 1982: 216) (emphasis is the author's). The social context of situation "speaks" in the sense that it indicates to the beginning learner, through its physical and human constitutive entities[15], which meanings are 'given' (those shared between the beginning learner and his/her

---

[12] Corder (1981a: 90) refers to the learner's IL continuum as "a dynamic, goal-oriented language system of increasing complexity".

[13] The same as simplification as I have interpreted these concepts. Jaworski's (1998a: 290) interpretation of Nemser – "Simplification and TRANSFER are the main features of learner talk described by Nemser (1974 [1971]) as *approximative system*" (emphases are the author's) – lends support to my own interpretation.

[14] Nemser (1974: 58) uses the term 'learner pidgin' to refer to the L_a of an advanced L2 learner who is fluent but has not mastered the L_t (L2) "... fundamentals ...".

[15] The human entities are: the actors / participants involved in the social interaction.

interlocutors) and which are 'new' (those that are not shared between the beginning learner and his/her interlocutors)[16]. That which makes it possible for the beginning learner to understand the context-of-situation's 'speaking' is his/her "... knowledge of the world [knowledge of the conceptual organization of events] and (...) of the informativeness principle in communication" (Ellis, 1982: 214), the second type of knowledge being a result from his/her previous experience with the L1. Once the 'given' and 'new' meanings have been identified within the physical and human dimensions of the context of situation, the beginning learner, in order to be able to codify, in the L2, the elements of his/her knowledge of the world he/she wishes to communicate, simplifies these elements through the elimination, from the message to be codified, of the 'given' meanings and through the emphasis on the 'new' meanings that are directly related to the here-and-now of the communicative situation. For Ellis, therefore, the 'initial IL', the first "mental grammar", the first "stable plateau" derives from a semantic simplification process. The subsequent IL stages, those which make up the continuum in the direction of the L2, are the product of a gradual complexification process of the initially simplified L2. Within the scope of this proposal, the complexification process is also susceptible, like in Selinker's, Corder's, and Nemser's proposals, to the effects of fossilization.

Hypothesis 4

> The L2 learner utilizes his capacity to learn, store, and reproduce verbal information to search for 'formulas' that will be communicatively useful for him (Ellis, 1982: 218) (the emphasis is the author's)[17].

indicates that the beginning learner increases the communicative efficiency of his/her messages in the 'initial IL' by way of the reproduction of formulas or fixed/stereotyped expressions that he/she spots in the interlocutors' output, memorizes, and then incorporates to his/her own output. Besides the interactional aspect of Hypothesis 4 as to the manner how the learner incorporates the formulas, its sociolinguistic dimension is also found in the fact that the higher or lower occurrence frequency of formulas in the 'initial IL' depends on two social factors: the type of learning environment (higher frequency in the classroom than in naturalistic environments); the degree of the communicative need (the greater the communicative need, the higher the occurrence of semantic simplification – the creative aspect of the 'initial IL' – and the lower the occurrence of formulas).

Hypotheses 2 and 3 are:

> The L2 learner knows that language is syntactic. He operates with the assumption that word order is meaningful if this is true for his L1 (Ellis, 1982: 216).

---

[16] For further information on the concepts of 'given', 'new', and 'information structure', the reader should refer to Chapter 8 (especially Sections 8.4, 8.5, and 8.6) in Halliday (1994).

[17] Corroborating Ellis's (1982) view, Myles *et al.* (1998: 324/327) claim, based on empirical evidence, that "... pre-fabricated routines and patterns, imitated utterances, formulas, formulaic units ..." display a high rate of occurrence in L2 classrooms, especially in the beginning stages. Moreover, these researchers state that the formulas' contribution to learning is twofold: initially, they help with the development of "... an emerging grammatical competence ..."; as the learning proceeds, they start being broken up or unpacked and their constituting parts give rise to new utterances, a process which leads to the formulation of new IL rules.

126

> The L2 learner knows that language realizes modality[18] elements as well as propositional elements and actively seeks out how to express those modal meanings that he considers communicatively useful (Ellis, 1982: 216).

The contribution these hypotheses bring to the emergence of the 'initial IL' is based on the cognitive dimension of the implicit or internalized knowledge that the L2 learner holds about the functioning of languages in general: that they are syntactic (word order has meaning), and that they express grammatical (modality) and content (propositional) meanings. At this point, I see a contradiction, on the part of Ellis, that impairs the internal coherence of his four hypotheses: whereas Hypothesis 1 states that the simplification strategy the beginning L2 learner resorts to is solely of the semantic type, Hypotheses 2 and 3 postulate that he/she makes use of word order and modality elements in a selective way, guided by communicative criteria. If the initial use of grammar is selective, it means that it is restricted and simplified. Accordingly, it is my claim that the beginning classroom L2 learner resorts to simplification that is both semantic and grammatical.

A more careful analysis of Ellis's (1982) four hypotheses leads to the conclusion that their strictly dichotomic separation into socially-natured and cognitively-natured hypotheses, as proposed by myself, is not exact. At least two of them display a mixed nature: the first one, here classified as mostly socially-oriented, depends on the knowledge of the world, which is a cognitive construct related to the semantic memory ("[...] our storage of generic information about the world and contains our knowledge about facts of nature, things we have learned at school, and our mental models of the world" [Fortkamp, 2000: 12]); the third one, here classified as mostly cognitively-oriented, depends, like the socially-natured Hypothesis 4, on the output of the interlocutors involved in the interaction, in order that the beginning learner can identify the structures that actualize the modal meanings he/she considers as relevant. Such hybrid character of two of Ellis's (1982) hypotheses corroborates the fact that, for him, the 'initial IL hypothesis' is not an all psycholinguistic issue, but it is also a sociolinguistic issue.

Ellis concludes his 'initial IL hypothesis' by proposing a three-stage developmental path for IL: "Stage 1: Semantic simplification + formulas [...] Stage 2: Semantic implementation + acquisition of **some** modality elements [...] Stage 3: Acquisition of **further** modality element (sic)" (1982: 220) (emphases are mine).

Ellis (1985: 252), reviewing Schumann (1978), says that the latter refers to his 'initial IL hypothesis' as the 'pidginization hypothesis', which is related to the socio-psychological process of acculturation. The acculturation process, in turn, is a theoretical construct that depends on the variables: 1) social distance – "[...] the result of [...] domination versus subordination, assimilation versus adaptation versus preservation, enclosure, size, congruence, and attitude" (McLaughlin, 1987: 110) of the foreign **group** (naturalistic L2 learners) in relation to the native group (L2 native speakers); 2) psychological distance – "[...] the result of [...] affective factors that concern the learner

---

[18] For Ellis, following Fillmore (1968), modality encompasses "[...] bound (e.g. -ed) and free (e.g. the definite and indefinite articles) grammatical functors" (Ellis, 1988: 172-173). For Halliday (1994: 88-92), on the other hand, modality is the lexico-grammatical means of modalizing (conveying degrees of probability and usuality) and modulating (conveying degrees of obligation and inclination) discourse.

as an **individual**, such as resolution of language shock, culture shock, and culture stress, integrative versus instrumental motivation, and ego permeability" (McLaughlin, 1987: 110) (emphasis is mine).

For Schumann's (1978, *apud* Ellis, 1985: 252) 'acculturation theory', then, the first learning stage of an L2 in a naturalistic setting follows the same processes as those that take place when a pidgin, in language contact situations, is emerging. In Cook's (1998: 1) interpretation: "The starting-point is the resemblance of pidgin languages to L2 learners' languages, particularly in terms of the overall simplicity of SYNTAX" (emphasis is the author's)[19]. The extent to which the first learning stage will develop, by means of its lexico-grammatical complexification, into subsequent stages is determined by the degree of social and psychological distances: the smaller the distances, the higher the probability of success; the greater the distances, the higher the probability of the first-learning-stage's remaining pidginized (McLaughlin [1987: 112] uses the term 'fossilized').

As reported by McLaughlin (1987: 112), Schumann's empirically-found lexico-grammatical features of the first L2 learning stage are:

(1)   Use of the general preverbal negators: 'no' 'don't'.
(2)   No question inversion.
(3)   Lack of auxiliary.
(4)   No inflection of possessive.
(5)   Use of uninflected forms of the verb.

I would say that these features are also present in the 'initial IL' of classroom learners of another language as a foreign – not as a second – language. Such a claim is based on my believing that the 'acculturation theory' applies to this group of learners as well: while the social distance is almost always maximal since these are not culturally-immersed learners, the psychological distance may assume the same dimensions as it does for naturalistic learners (e.g.: a Brazilian whose political convictions are strongly left-winged sits in an EFL classroom solely due to educational and/or professional demands).

Building upon Schumann's 'acculturation theory', with a greater emphasis on the cognitive dimension, Andersen (1983, *apud* Ellis, 1985; McLaughlin, 1987) creates his 'nativization theory' around the processes of 'nativization' (equivalent to the Piagetian concept of assimilation) – responsible for the 'initial IL' – and 'denativization' (equivalent to the Piagetian concept of accommodation) – responsible for the subsequent learning stages.

At the beginning, the learner builds an L2 system that is 'native' to him/her in the sense that it is autonomous in relation to the L2 norm and, thus, unique. The building of such a unique system depends upon the L2 input the learner is exposed to and is guided by "... his own internalized view of what constitutes the L2 system" (Ellis, 1985: 253). There is, therefore, due to a "... relatively restricted [initial] access to target-language [L2] input" (McLaughlin: 1987: 113) in naturalistic settings, the action of a simplification strategy in that the basis for the construction of the L2 knowledge is the learner's previous

---

[19] Jaworski (1998b: 248) states that the simplification of pidgins encompasses not only the level of syntax but also the levels of lexis and morphology.

knowledge of both the L1 and the world and his/her perceived knowledge of the L2[20].
Andersen (1981, *apud* McLaughlin, 1987: 117), based on the comparison between a
pidgin language and the production of a beginning L2 learner, came up with the following
effects of the simplification strategy:

(1) reliance on word order rather than inflections for expressing grammatical relations.
(2) native-language transfer in word order as well as use of English word order.
(3) sporadic emergence of preverbal markers which come from lexical verbs promoted to auxiliary status.
(4) a basic pidgin negation.
(5) lack of inversion in questions.
(6) preponderance of uninflected verb forms.

Throughout the following stages, during which the exposure to L2 input tends to
increase, the uniqueness of the learner's initial L2 system tends to disappear, i.e., it
undergoes the process of denativization. It means that the learner's initially simplified
L2 system becomes, by way of its complexification, ever more similar to the L2 input.

Between the exclusively psycholinguistic position and the mixed
socio/psycholinguistic one, my theoretical choice is for the latter. Such an option is
motivated by my belief in the fact that an L2 learner cannot be considered as an isolated
'cognition island' (intrapersonal dimension). On the contrary, I support the idea
whereby the learner's cognitive apparatus relates itself bidirectionally both with the
physical world and with the other actors that participate in the acquisition process
through interaction and are themselves holders of cognitive apparatuses. The other
actors are teachers, colleagues, and native and/or non-native interlocutors in general
(interpersonal dimension).

Among the representatives of the socio-psycholinguistic position reviewed in this
section, my choice is for the 'initial IL hypothesis' raised by Ellis (1982) as it is the
broadest in scope and the only one to have been built directly upon Selinker's (1974)
original 'IL theory'. However, its scope still needs broadening through the incorporation
of structural simplification, besides semantic simplification, in the terms of the early
developmental stages Ellis himself proposes for naturalistic learners: "[...] *a silent period*
[...] the use of *formulaic speech*, and [...] *structural and semantic simplification*" (1994:
82) (emphases are the author's).

My choice for Ellis's (1982) 'initial IL hypothesis', with the incorporation of
structural simplification to semantic simplification, justifies my methodological choice
for SFG as a data categorization framework[21]. The reasons are the following: 1) a
cognitive-and-social 'IL theory' is better accompanied, as it were, by a socially-
grounded linguistic theory than by a non-social one like Fillmore's (1968) Case
Grammar, which was used by Ellis (1988); 2) in SFG, grammar encompasses syntax,
morphology, and vocabulary, thus being called lexico-grammar (Halliday, 1994: xiv);
3) the structures (lexico-grammatical stratum) are the motivated/natural actualizations

---

[20] This aspect of the 'nativization theory' has much in common with Ellis's (1982) Hypothesis 1.

[21] My choice is further justified by what Perrett (2000: 89) says: "The theoretical framework of SFL
[Systemic-Functional Linguistics] and some of its practical procedures can provide particular advantages for
the study of SLD [Second Language Development], especially with regard to how language use changes over
time".

of the metafunctions (semantic stratum), which means that, insofar as SFG is concerned, semantics and grammar are inseparable – the borderline between the two is almost non-existent; 4) in SFG terms, Ellis's structural and semantic simplification can be renamed as lexico-grammatical simplification.

2. METHODOLOGY

2.1 Subjects

Ten Brazilian EFL students from the *CCB/UFC*, located in Fortaleza, Ceará, took part in this study. The candidates had to meet the following minimum conditions: (1) age range – from 20 to 35; (2) level of instruction – to be at higher education[22]; (3) level of English proficiency – beginning.

In order to control for the level-of-English-proficiency variable, the potential subjects were those enrolled, at data collection time – January/2002, in 'Semester 2' (S2), the second stage of the beginning level within the *CCB/UFC* seven-semester EFL program[23]. The need to be consistent, as much as possible, with Studies 1 and 2, which adopted the traditional three-level criterion of proficiency classification (beginning, intermediate, advanced), led me to consider the first third of the *CCB/UFC* program (S1, S2, and the first half of S3) as its beginning level, the second third (the second half of S3, S4, and S5) as its intermediate level, and the last third (S6 and S7) as its advanced level. I did not choose to work with either S1 or the first half of S3 due to the methodological recommendation, made in Study 2 (cf. V 40), to use the middle stage of a given proficiency level for a more accurate 'frozen picture' of that level in cross-sectional studies. Such a recommendation was based on the results presented in the aforementioned study.

The ten subjects were randomly chosen. When the data were collected, the *CCB/UFC* EFL program had twelve S2 classes, taking place at different times. As a first step, seven of the classes were randomly eliminated. The second step consisted of my going, immediately before data collection, into the remaining five classes and asking, in each of them, for two volunteers who were university students at the age of 20 or over.

2.2 Corpus

The corpus is made up of pairs of spoken/written texts (twenty texts altogether) that were rendered, in an impromptu manner, by each of the ten subjects within the narrative rhetorical mode. Both the spoken and the written narratives belong to the genre of remarkable personal experiences.

---

[22] The justification for conditions (1) and (2) as well as for the total number of subjects – 10 – derives from a personal interest to make the results of the present study comparable with those of Study 1, with advanced EFL students, and Study 2, with intermediate EFL students. The 10 students who were available to be the subjects of Study 1 happened to hold the characteristics expressed in (1) and (2).

[23] Each stage lasts for one academic semester, and the adopted series of books is *English File* (Oxenden and Seligson, 1996), which is a three-volume series. The S2 syllabus covers lessons 4, 5, and 6 of the first volume.

2.3 Data collection procedures

After receiving the consent of the coordinator of the *CCB/UFC*, I contacted the five teachers of each one of the selected S2 classes for their permission. With each teacher, the day for my visit was then set up, which was the same for all classes.

On the appointed day and at the beginning of each class, I entered the classroom, together with the respective teacher. The teacher introduced me to the students, and I told them I was a researcher within the SLA area and needed two twenty-year-old (or over) university students as volunteers to do some simple tasks whose outcomes would be used as data for the investigation I was carrying out on how Brazilians learned EFL. Since the coordinator had made available a room that was located close to the rooms where the previously selected S2 classes were held, I invited one of the volunteers to follow me there. Once that subject was finished with the tasks, I would accompany him/her from the data-collection room to his/her room. As soon as the first volunteer was back, the same was done with the second one.

In the data-collection room, the subject and myself would sit at opposite sides of a table. I then asked him/her, in Portuguese[24], to tell me, in a monologic way[25] and in English, a remarkable experience, positive or negative, he/she had lived through. Before the subject began speaking, I asked him/her whether I would have the permission to audio tape the monologic narrative (all of the 10 subjects granted the permission for the recording). The subject was allowed to speak for as long as he/she wanted. When the subject had finished, he/she was immediately given a sheet of paper, a pencil, and an eraser and was requested (in Portuguese) to tell, in writing (in English), the same remarkable experience. The writing time was not controlled either. Each data-collection session lasted, on the average, for 15 minutes.

The data-collection conditions made it possible for the narratives to be rendered in an impromptu manner as all of the subjects walked into the data-collection room without knowing any information about what the tasks were like. The subjects had very little planning time for the production of the two narratives.

When the subjects had finished with the two tasks, I asked them (also in Portuguese): 1) whether they had found the spoken task difficult; 2) what they had found most difficult on doing the spoken task; 3) which task – the spoken one or the written one – they had found more difficult and why; 4) whether they had entered *CCB/UFC* in S1 or whether they had taken a placement test; 5) what their previous experience with English had been, if any. To three of the subjects, it was necessary to ask for clarification as to the meaning of some parts of their spoken narratives.

2.4 Data categorization procedures and data analysis criteria

The first categorization procedure was the transcription, into orthographic script, of the spoken narratives. There are, however, some differences between the adopted orthographic script and the regular orthographic script utilized in the graphic linguistic

---

[24] The choice of the Portuguese language as a means of instruction was due to the subjects' being beginning learners.

[25] The subjects were notified that no questions could be asked and that they had to act as if I were not in the room.

channel: the former lacks paragraph indentation, capital letters, and punctuation marks. The reason for this lies in the fact that supra-segmental phonological aspects are outside the scope of the present study.

The 20 narratives were segmented into both ranking and down-ranked clauses (cf. the Appendix for samples of transcribed and categorized narratives). Their structural constituents were categorized as for Halliday's (1994) configurational functions which realize the systems of transitivity (that instantiates the ideational metafunction) and mood (that actualizes the interpersonal metafunction). The clause segmentation criterion was a semantic-structural (grammatical) one. The departure point, chosen as ideal for the identification of clause extension, was the Process, due to the fact that, within the transitivity configuration, both the Participants and the Circumstances agglutinate around it. When there is a single lexical verb, the Process is realized by a simple verbal group and, hence, there is only one clause. However, when there are two adjacent lexical verbs, the following situations are possible: 1) the two lexical verbs make up a single Process that indicates the existence of only one finite clause as the Process is realized by a verbal group complex; 2) each of the two lexical verbs makes up a different Process, which means that there are two simple verbal groups (not one verbal group complex) and, thus, two distinct clauses – the first is finite, and the second, non-finite (for lack of space, it is not possible to provide further details; the reader, however, should refer to Halliday's [1994] Chapter 7).

The ranking clauses were quantified separately for the spoken and written narratives. The complete ranking clauses for transitivity and mood and the incomplete ones for transitivity or mood (altogether and per type: LAP, LPP, SDCC, and the respective down-ranked-clause-related incompletenesses) were also quantified per medium. As the narratives have different lengths, the absolute numbers that resulted from these countings were turned, for the sake of overcoming the length problem, into simple frequency indices as proposed by Beaman (1984), which, in turn, were transformed into percentages. Examples, from the data, of each clause category are as follows:

- Complete Ranking Clauses (CC):

Spoken Narrative 01/Clause 1

	I	like		my family
TRANSITIVITY	Senser	Process: Mental		Phenomenon
MOOD		present	like	
	Subject	Finite	Predicator	Complement
	Mood		Residue	

Written Narrative 01/Clause 4

	My house	is		big and confortable (sic)
TRANSITIVITY	Carrier	Process: Relational		Attribute
MOOD		present	be	
	Subject	Finite	Predicator	Complement
	Mood		Residue	

- LAP Incomplete Ranking Clauses – Actual Incompletenesses (IC-LAP):

2‖ I haved one one grand grandmother 3‖ **and I like very very much 4** ‖ and she die ‖

Spoken Narrative 03/Clause 3

	and	I	like		Φ ·	very very much
TRANSITIVITY	X	Senser	Process: Mental		Phenomenon lacking	Circumstance
MOOD	X		past (Φ)	like		
	X	Subject	Finite lacking	Predicator	Complement lacking	Adjunct
	X	incomplete Mood			incomplete Residue	

Written Narrative 03/Clause 4

	but	Φ	was		very dificult (sic)	for me
TRANSITIVITY	X	Carrier lacking	Process: Relational		Attribute	Circumstance
MOOD	X		past	be		
	X	Subject lacking	Finite	Predicator	Complement	Adjunct
	X	incomplete Mood			Residue	

- LAP Incomplete Ranking Clauses (down-ranked-clause-related-incompleteness):
  - ➢ No occurrences of this category in the data.
- LPP Incomplete Ranking Clauses – Pseudo Incompletenesses (IC-LPP):

1‖ **in last year I 2**‖ I traveled with friends by plane ‖

Spoken Narrative 02/Clause 1

	in last year	I	Φ		Φ	Φ
TRANSITIVITY	Circumstance	Actor	Process: Material lacking		Circumstance lacking	Circumstance lacking
MOOD			past (Φ)	Φ		
	Adjunct	Subject	Finite lacking	Predicator lacking	Adjunct lacking	Adjunct lacking
	incomplete...	incomplete Mood			...Residue	

- LPP Incomplete Ranking Clauses (down-ranked-clause-related-incompleteness):
  - ➢ No occurrences of this category in the data.
- SDCC-Related Incompletenesses:

2‖ when I arrived 3‖ **I see [[my dog die]]** ‖

Spoken Narrative 10/Clause 3

	I	see		[[my dog die]]
TRANSITIVITY	Senser	Process: Mental		Phenomenon
MOOD		present	see	
	Subject	simplified deviant Finite [simple present for simple past]	Predicator	Complement
	simplified deviant Mood		Residue	

133

8 ‖ I cryed (sic) 9 ‖ **and runned to the telephone** (Anaphoric Elliptical Clause-AEC) ‖

Written Narrative 10/Clause 9

	and	*I*	runned		to the telephone
TRANSITIVITY	X	*anaphorically retrieved Actor*	Process: Material		Circumstance
MOOD	X		past	run	
	X	*anaphorically retrieved Subject*	simplified deviant Finite [deviant past marking → overgeneralization]	Predicator	Adjunct
	X	simplified deviant Mood			Residue

- SDCC-Related Incompletenesses (down-ranked-clause-related-incompleteness):
  - ➢ No occurrences of this category in the data.

The determination of the level (low, moderate, or high) of lexico-grammatical complexification of the subjects' IL as represented by their narratives is based on the following *a priori* criteria: up to 50% of completeness – low complexification level; from 50% (exclusive) to 80% (inclusive) of completeness – moderate level of complexification; greater than 80% of completeness – high level of complexification. These criteria, thus, aim at the verification of the hypotheses as an outcome of the analysis of the data.

3. DATA ANALYSIS

The simple frequency index (SFI), indicative of the occurrence rate of a certain feature in a group of texts per every 1,000 words, is the result of the ratio between the absolute number of occurrences of that feature and the total number of words of the group of texts. The resulting ratio is then multiplied by 1,000. Hence, the total words in the spoken and written narratives were quantified. The quantifications are displayed in Table 1.

	Spoken Narratives	Written Narratives
Total Numbers of Words	727	477

Table 1: Total numbers of words in the narratives per medium

As noted in the 'Introduction', one of the contributions of Study 2 (cf. V 40) was the creation of the SDCC category. Aiming at testing its soundness, the hypotheses of the current study will be verified from the perspectives of the non-incorporation and of the incorporation of the innovative category.

3.1 Hypothesis 1: [Since the subjects are beginning EFL learners, their spoken and written narratives (both separately and together) bear a low level of lexico-grammatical complexification]

134

- Perspective 1:

The results for the spoken and written narratives separately are shown in Table 2.

	Complete Ranking Clauses + Total Pseudo-Incompletenesses [LPP Incomplete Ranking Clauses + LPP Incomplete Ranking Clauses (down-ranked-clause-related)]	Total Actual Incompletenesses [LAP Incomplete Ranking Clauses + LAP Incomplete Ranking Clauses (down-ranked-clause-related)]	Total Ranking Clauses
Spoken Narratives	107.29 (50.98%)	103.16 (49.02%)	210.45 (100%)
Written Narratives	117.40 (66.67%)	58.70 (33.33%)	176.10 (100%)

Table 2: Simple frequency indices and respective percentages for complete ranking clauses, incomplete ranking clauses, and total ranking clauses in the spoken & written narratives separately

It is evidenced in Table 2 that the spoken narratives contain about as many complete and total pseudo-incomplete ranking clauses as total incomplete ranking clauses. This means, for the spoken medium, a complexification level of 50.98% and a simplification level of 49.02%. The written narratives, in turn, contain about twice as many complete and total pseudo-incomplete ranking clauses as total incomplete ranking clauses. This means, for the written medium, a complexification level of 66.67% and a simplification level of 33.33%.

Table 3 brings the results for the spoken and written narratives together.

	Complete Ranking Clauses + Total Pseudo-Incompletenesses [LPP Incomplete Ranking Clauses + LPP Incomplete Ranking Clauses (down-ranked-clause-related)]	Total Actual Incompletenesses [LAP Incomplete Ranking Clauses + LAP Incomplete Ranking Clauses (down-ranked-clause-related)]	Total Ranking Clauses
Spoken & Written Narratives	224.69 (58.13%)	161.86 (41.87%)	386.55 (100%)

Table 3: Simple frequency indices and respective percentages for complete ranking clauses, incomplete ranking clauses, and total ranking clauses in the spoken & written narratives together

As can be seen in Table 3, the spoken and written narratives encompass, together, around 1.4 as many complete and total pseudo-incomplete ranking clauses as total incomplete ranking clauses. This means complexification and simplification levels, regardless of medium, of 58.13% and 41.87%, respectively.

Insofar as Perspective 1 is concerned, Hypothesis 1 was refuted. For the spoken narratives separately, the refutation, however, seems not to have been relevant due to the small percentage margin of 0.98%. For the written narratives separately and the spoken and written narratives together, on the other hand, the refutation seems to be relevant as the percentage margins are 16.67% and 8.13%, respectively.

- Perspective 2:

Tables 4 and 5 display the results when the (SDCC + SDCC/down-ranked-clause-related) type of incompleteness is incorporated to the total incomplete ranking clauses.

135

The former contains the Perspective2-results for the spoken and written narratives separately.

	Complete Ranking Clauses + Total Pseudo-Incompletenesses [LPP Incomplete Ranking Clauses + LPP Incomplete Ranking Clauses (down-ranked-clause-related)]	Total Actual Incompletenesses [LAP Incomplete Ranking Clauses + LAP Incomplete Ranking Clauses (down-ranked-clause-related)] + [SDCC + SDCC (down-ranked-clause-related)]	Total Ranking Clauses
Spoken Narratives	96.28 (45.75%)	114.17 (54.25%)	210.45 (100%)
Written Narratives	96.44 (54.76%)	79.66 (45.24%)	176.10 (100%)

Table 4: Simple frequency indices and respective percentages for complete ranking clauses, incomplete ranking clauses + the SDCC category, and total ranking clauses in the spoken & written narratives separately

Table 4 demonstrates that the spoken narratives have around 1.2 times as many total incomplete ranking clauses with the total SDCC category incorporated as complete and total pseudo-incomplete ranking clauses, which means, for the spoken medium, a complexification level of 45.75% and a simplification level of 54.25%. The written narratives, on the other hand, bear 1.2 times as many complete and total pseudo-incomplete ranking clauses as total incomplete ranking clauses with the total SDCC category incorporated, which means, for the written medium, a complexification level of 54.76% and a simplification level of 45.24%.

In Table 5, the Perspective2-results for the spoken and written narratives together can be found.

	Complete Ranking Clauses + Total Pseudo-Incompletenesses [LPP Incomplete Ranking Clauses + LPP Incomplete Ranking Clauses (down-ranked-clause-related)]	Total Actual Incompletenesses [LAP Incomplete Ranking Clauses + LAP Incomplete Ranking Clauses (down-ranked-clause-related)] + [SDCC + SDCC (down-ranked-	Total Ranking Clauses
Spoken & Written Narratives	192.72 (49.86%)	193.83 (50.14%)	386.55 (100%)

Table 5: Simple frequency indices and respective percentages for complete ranking clauses, incomplete ranking clauses + the SDCC category, and total ranking clauses in the spoken & written narratives together

As evidenced in Table 5, the spoken and written narratives, when considered together, present about as many complete and total pseudo-incomplete ranking clauses as total incomplete ranking clauses with the total SDCC category incorporated, which means,

independently of the medium, a complexification level of 49.86% and a simplification level of 50.14%.

Inasmuch as Perspective 2 is concerned, whereas Hypothesis 1 was confirmed as regards the separate spoken narratives and the spoken and written narratives together, it was refuted concerning the separate written narratives. The refutation, however, was by a percentage margin of 4.76%, which seems to be irrelevant.

3.2 Hypothesis 2: [Since the writer has more time than the speaker to elaborate on his/her discoursal production, the subjects' written narratives will bear a higher level of lexico-grammatical complexification than their spoken ones]

- Perspective 1:

In accordance with Table 2, the complexification levels of the spoken and written narratives taken separately are, respectively, 50.98% and 66.67%. The difference of 15.69% in favor of the written medium leads to the confirmation of Hypothesis 2.

- Perspective 2:

Table 4 indicates that the complexification level of the separate spoken narratives is 45.75% and of the separate written ones is 54.76%. The difference of 9.01%, towards the written medium, also confirms Hypothesis 2.

Compared to Studies 1 and 2, these Hypothesis2-confirmation figures (15.69% and 9.01%) for the beginning level of EFL instruction are larger. While Study 1 confirmed Hypothesis 2, for the advanced level, by a difference of 6.53%, Study 2 (cf. V 40) confirmed it, for the intermediate level, with the figures 2.72% (Perspective 1: non-incorporation of the down-ranked-clause-related incompletenesses) and 1.39% (Perspective 2: incorporation of the down-ranked-clause-related incompletenesses).

In both Study 2 (cf. V 40) and the present paper, each respective Perspective 2 is the criterion for the confirmation/refutation of the hypotheses.

Four aspects of the results presented in the above analysis will be discussed in search of plausible interpretations: 1) the refutation of Hypothesis 1 as for the separate written narratives; 2) the borderline situation (between low and moderate lexico-grammatical complexification levels) of the figures indicative of the level of complexification for the three research conditions; 3) the larger figure through which Hypothesis 2 was confirmed; 4) the non-occurrence of down-ranked-related-incompletenesses in the beginning level.

4. DISCUSSION

4.1 Refutation of hypothesis 1 as for the separate written narratives

A similar result was found, for intermediate learners, in Study 2 (cf. V 40). In that study, I claimed that this result was not surprising, that it would have been unexpected if it had been the contrary: refutation as for the spoken narratives and confirmation as for the written ones. Furthermore, I postulated that a possible explanation for the phenomenon would be the assumption part of Hypothesis 2: the writer has more time than the speaker to elaborate on his/her discoursal production.

In response to the *a posteriori* question as to which task – the spoken one or the written one – they had found more difficult and why (Question 3 of the short interview I conducted, in Portuguese, with the subjects), five of the subjects corroborated the assumption part of Hypothesis 2, with the following spontaneous comments:

SUBJECT	RESPONSES
Sub 5	"I think what is more difficult is to structure the phrases in order to speak. When writing, you think longer. I think you start to write the phrases, see what isn't right, cross it out, and do it all over again; when speaking, you can't do that. I think longer when I'm writing; I think this is a difference between speech and writing. In speech, you have to think and speak almost automatically; otherwise, the speech fluency decreases a lot. In writing, you can think longer and elaborate longer in order to be able to write".
Sub 6	"The spoken task because, when writing, we have more … we memorize certain words. On the contrary, when speaking we haven't had contact with all of the words. Generally in writing, we have at least seen them; we just don't know their pronunciation; in writing,, we memorize them in a certain way. In oral language, you feel inhibited to speak some words you had contact only visually because you don't know the correct pronunciation".
Sub 7	"The spoken task because, for a beginner, to elaborate the ideas and remember all the vocabulary, and to coordinate, I find it difficult. I think speaking is more difficult than writing. When writing, you are more at ease.
Sub 9	"I think that, in writing, you've got more time to think; in my case, to think in Portuguese and translate when I'm writing. Speech flows faster, and we don't have time to correct ourselves".
Sub 10	"The spoken task because speech is almost instantaneous. You've got to throw the ideas quickly, converting them into English. In writing, I've got more time, more tranquility to convert".

Figure 3: Subjects' responses to which task presented more difficulty

These comments seem to point to the direction that the assumption part of Hypothesis 2 (the on-line character of speech and the planning character of writing, due to the existence or not of time restrictions, make a difference in the lexico-grammatical complexification level of L2 learners' output) is a sound explanation for the result found both here and in Study 2 (cf. V 40). An authoritative support for such a claim comes from Widdowson (1979: 194/208-209):

> […] it is not uncommon to find lapses which occur regularly in speech but not at all, or not regularly, in writing […] literacy provides conditions which favour the development of babu [surface elaboration and verbosity as opposed to the simplification of pidgins]. Written language is of its nature independent of immediate context and something, therefore, that can be fashioned in detachment as an artefact.

4.2 The borderline situation between low and moderate complexification levels

This aspect in itself is another possible explanation for the first one. It is true that Hypothesis 1 was refuted as to the separate written narratives; however, Tables 4 and 5 show that the levels of lexico-grammatical complexification for the separate spoken narratives and for the spoken and written ones considered together – 45.75% and 49.86%, respectively – are not very far from that of the separate written narratives (54.76%), within the range 45%-55%.

These figures evidence: 1) that, for the research conditions 1 and 3 (separate spoken narratives and spoken and written narratives together, respectively), the subjects bear low lexico-grammatical levels of complexification, but they are very close to the moderate

level; 2) that, for the research condition 2 (separate written narratives), the subjects are already within the moderate lexico-grammatical level of complexification, but, on the contrary, they are not yet very far from the low level.

These results might be indicative of the fact that the S2 subjects, even though still in the beginning instruction level in accordance with the traditional three-level classification, must have already gone through the 'initial IL' stage, the first "mental grammar", the first "stable plateau" of the "built-in syllabus" and be somewhere beyond already, especially due to the claims put forward by Larsen-Freeman and Long (1991: 82):

> First, compared with some other forms of language growth, ILs are typically changing rather fast in developmental terms. One reason for this is that most learners' earliest IL grammars are soon being modified towards an accessible, external, target-language [L2] norm, a process Andersen (1979b, p. 109) refers to as 'denativization'. Availability of the target-language model means this often happens relatively quickly [...] A second reason for the rapid developmental change is that older children and adult SL [Second Language] acquirers are less cognitively and psycholinguistically constrained than young children acquiring their native language [...] This means that, while learners in early stages may, as Andersen (1979b, p. 109) claims, be creating their own internal representation of the SL, a composite of processes he calls 'nativization', even early SLA tends to proceed rather fast.

Therefore, insofar as Ellis's (1982) three-stage IL developmental proposal is concerned, the S2 subjects must have surpassed Stage 1 and must be in Stage 2.

Besides the explanatory claims suggested by Larsen-Freeman and Long (1991), there are two other possible and concomitant explanations. The first has to do with the narrative rhetorical mode of the elicited data. Doughty and Long (2000: 152), reviewing research done about the effects – on L2 accuracy, fluency, and grammatical complexity – of the dependence or not of the narratives' topic on the here-and-now, state that the reported findings indicate "[...] that an orientation to displaced time and space[26] results in trends toward lesser fluency but greater accuracy and complexity of output". The second regards the subjects' previous experience with English: four of them entered *CCB/UFC* in S2 through a placement test (Sub 2 had had a year's study experience; Sub 3 and Sub 4, a six months' experience; Sub 9, a two months' experience); out of the six of them who started in *CCB/UFC*'s S1, two had studied previously at private Language Centers and four had had a restricted-to-secondary-school-instruction experience.

4.3 The more effective confirmation of Hypothesis 2

In this study, Hypothesis 2 was more soundly confirmed (9.01%) than in Studies 1 and 2. In the latter study, I raised the following as a likely explanation: the small confirmation margins (1.39% and 6.53%, for advanced and intermediate learners, respectively) might be an indication that the lexico-grammatical complexification paces are almost the same for both the spoken and the written IL media.

The new evidences indicate that, if the aforementioned explanation is relevant for the advanced and intermediate levels of instruction, it might not be true for the

---

[26] Dependence on the there-and-then (displaced time and space) is the characteristic of the narratives elicited for this study.

beginning one. This claim is in tune with what Larsen-Freeman and Long (1991: 82) posit as to the rapidity with which development unfolds for beginning L2 learners.

4.4 The non-occurrence of down-ranked-related-incompletenesses

As a matter of fact, the occurrence rate of down-ranked clauses in general, even those that are not related to the incompleteness of outer or ranking clauses (the down-ranked-related-incompletenesses), was very low: in absolute numbers, only three in the spoken narratives and only one in the written narratives. They are: Spoken Narrative 06 / Clause 9: ‖ we see churches [[on which ‖ on which have beautiful beautiful *imagens ícones*]] ‖ ; Spoken Narrative 10 / Clause 3: ‖ I see [[my dog die]] ‖ ; Written Narrative 10 / Clause 7: ‖ At home I saw [[my dog die.]] ‖ .

My interpretation for this phenomenon is that, although the S2 subjects might have already gone through the 'initial IL' stage, they are still within the beginning instruction level as regards the traditional three-level classification. This is so because, in my view, a very low occurrence of embedding or down-ranking is a very strong indication of lexico-grammatical simplicity.

Another indication of lexico-grammatical simplicity, according to Praxedes Filho (1996) – based, in turn, on Halliday (1989), is the low occurrence level of clause complexes. In relation to the present study, its corpus as a whole (spoken and written narratives together) presents an SFI of 45.68 for the total occurrences, per every 1,000 words, of connectors, out of which 4.98 are of the hypotactic type (the conjunctions 'because' and 'when' and the relative pronoun 'which'), and 40.70 are of the paratactic type (the conjunctions 'and' and 'but'). Even without any comparison parameter, it seems to me that these frequencies are low.

Accordingly, it appears to be a sensible claim the suggestion that the simplification/complexification definition of learners' ILs used thus far ought to be expanded. It should be a multi-dimension definition that would encompass three continua: 1) the transitivity-and-mood functional-configurational continuum, with the incorporation of the categories of LAP, LPP, SDCC, and down-ranked-related incompletenesses; 2) a clause-complex continuum (simple clauses vs. clause complexes: from low to high occurrence of clause complexes)[27]; 3) a down-ranked-clause continuum (non-embedding vs. embedding: from low to high occurrence of down-ranked clauses).

5. CONCLUSION

The objectives initially set for this investigation were fully met, and the hypotheses raised about beginning learners were thoroughly verified. Both hypotheses were confirmed, except for the second condition of Hypothesis 1 – the complexification level of the separate written narratives, but the refutation seems not to bear much relevance due to the small margin (4.76%). The confirmation of the same hypotheses in Study 1, for

---

[27] If the SLA 'acculturation theory' postulates that the 'initial IL' is a pidgin-like language, a good support for this suggestion is what Romaine (1994: 174), quoted by Jaworski (1998b: 288), says about pidgin grammars: "The grammar of pidgin languages is 'shallow', for example, with no syntactic marking of subordination" (emphasis is the author's).

advanced learners, and in Study 2 (cf. V 40), for intermediate learners, leads me to claim that SFG has definitely proven to be an adequate data-categorization framework to be utilized in a longitudinal study that aims at researching the simplification-complexification continuum of Brazilian EFL learners' "built-in syllabus".

The second condition of Hypothesis 1 would most likely have been confirmed if the chosen *CCB/UFC* stage had been S1 as the results indicate that the S2 subjects who were informants in the present study might be on the verge of becoming intermediate learners, considering the traditional three-level criterion of proficiency classification. In a likewise manner, the results in Study 2 (cf. V 40) indicate that the top-intermediate stage subjects are on the verge of becoming advanced learners. As a follow-up, it appears that, for cross-sectional IL developmental studies, the most prototypical stages, again considering the traditional three-level classification, are: 1) the very first stage within the beginning level; 2) the middle stage within the intermediate level; 3) the top stage within the advanced level.

The fact that it was demonstrated that the beginning S2 subjects have already come out of the 'initial IL' stage is an evidence of the inadequateness of the traditional three-level classification as it does not reflect the learners' "built-in syllabus". For the proposal of a more realistic criterion of proficiency level classification, it is necessary that the boundaries of each IL stage along the simplification-complexification continuum be determined. However, only a longitudinal study is capable of reaching such an aim.

What IL stage, if not the initial one any longer, the S2 subjects utilized in this study are in? Pedagogically speaking, which is a more suitable syllabus to be adopted in the EFL classroom: that which is proposed by the methodologist and thus external to the learner, or that which is found out by means of an investigation into the IL lexico-grammatical simplification-complexification continuum possibly capable of evidencing the "built-in syllabus" for a given group of learners, which, in turn, could be used to teach subsequent groups? The answers to these questions also depend on longitudinal investigations.

The discussion of Aspect 2 led to the conclusion that the subjects must have surpassed Ellis's (1982) developmental Stage 1. His three-stage IL developmental proposal, based on his four 'initial IL hypotheses', uses Case Grammar as the theoretical foundation. A new IL developmental framework that is fully SFG-oriented ought to be proposed. Again, such a proposal cannot be dissociated from the longitudinal type of research.

However, before engaging into studying SLA longitudinally, I still need to carry out one more cross-sectional study, with the following characteristics: 1) the three instruction levels should be investigated at the same time and at the same teaching institution; 2) different types of EFL teaching institutions should be investigated simultaneously; 3) the narrative rhetorical mode should not be the only one used for data elicitation; 4) non-elicited, naturalistic data should also be analyzed; 5) the expansion of the simplification/complexification definition, suggested in 5.4, by the addition of the clause-complex continuum and the down-ranked-clause continuum should be tested.

REFERENCES

BEAMAN, K. (1984). Coordination and subordination revisited: syntactic complexity in spoken and written narrative discourse. In: TANNEN, D. (Ed.). *Coherence in spoken and written discourse*. Norwood, NJ: Ablex, pp. 45-80.

COOK, V. (1998). Acculturation hypothesis. In: JOHNSON, K.; JOHNSON, H. (Eds.). *Encyclopedic dictionary of Applied Linguistics*: a handbook for language teaching. Oxford: Blackwell, p. 1-2.

CORDER, S. P. (1978). Language-learner language. In: RICHARDS, J. C. (Ed.). *Understanding second & foreign language learning*. Rowley, Mass: Newbury House, pp. 71-93.
_____. (1981a). *Error analysis and interlanguage*. Oxford: Oxford University Press, 120p.
_____. (1981b). Formal simplicity and functional simplification in second language acquisition. In: ANDERSEN, R. W. (Ed.). *New dimensions in second language acquisition research*. Rowley, Mass: Newbury House, pp. 146-152.
_____. (1992). A Role for the mother tongue. In: GASS, S.; SELINKER, L. (Eds.). *Language transfer in language learning*. Amsterdam and Philadelphia: John Benjamins, pp. 18-31.

DOUGHTY, C.; LONG, M. H. Eliciting second language speech data. In: MENN, L.; RATNER, N. B. (Eds.). *Methods for studying language production*. Mahwah, NJ and London: Lawrence Erlbaum, 2000. p. 149-177.

ELLIS, R. (1982).The Origins of interlanguage. *Applied Linguistics*, Oxford, v. 3, n. 3, pp. 207-223.
_____. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press, 327 p.
_____. (1988). *Classroom second language development*: a study of classroom interaction and language acquisition. London: Prentice Hall, 233 p.
_____. (1994). *The Study of second language acquisition*. Oxford: Oxford University Press, 824 p.
_____. (1997). *Second language acquisition*. Oxford: Oxford University Press, 147 p.

FILLMORE, C. H. (1968). The Case for case. In: BACH, E.; HARMS, R. T. (Eds.). *Universals in linguistic theory*. New York: Holt, Rinehart and Winston, pp. 1-88.

FORTKAMP, M. B. M. (2000). Working memory capacity and L2 speech production: an exploratory study. Florianópolis, 300 f. Tese (Doutorado em Inglês e Lingüística Aplicada) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina.

HALLIDAY, M. A. K. (1989). *Spoken and written language*. 2. ed. Oxford: Oxford University Press, 109 p.
_____. (1994). *An Introduction to Functional Grammar*. 2. ed. London: Edward Arnold, 434 p.

HAWKINS, R. (1998). Error analysis. In: JOHNSON, K.; JOHNSON, H. (Eds.). *Encyclopedic dictionary of Applied Linguistics*: a handbook for language teaching. Oxford: Blackwell, pp. 110-114.

JAWORSKI, A. (1998a). Simplified codes. In: JOHNSON, K.; JOHNSON, H. (Eds.). *Encyclopedic dictionary of Applied Linguistics*: a handbook for language teaching. Oxford: Blackwell, pp. 288-293.
_____. (1998b). Pidgins and creoles. In: JOHNSON, K.; JOHNSON, H. (Eds.). *Encyclopedic dictionary of Applied Linguistics*: a handbook for language teaching. Oxford: Blackwell, p. 248.

JOHNSON, H. (1992). Defossilizing. *ELT Journal*, Oxford, v. 46, n. 2, pp. 180-189.

JOHNSON, K. (1998). Communication strategies. In: JOHNSON, K.; JOHNSON, H. (Eds.). *Encyclopedic dictionary of Applied Linguistics*: a handbook for language teaching. Oxford: Blackwell, pp. 57-62.

LARSEN-FREEMAN, D.; LONG, M. H. (1991). *An Introduction to second language acquisition research*. London and New York: Longman, 398 p.

McLAUGHLIN, B. (1985). *Second-language acquisition in childhood*: volume 2. school-age children. 2. ed. Hillsdale, NJ: LEA, 288 p.

McLAUGHLIN, B. (1987). *Theories of second-language learning*. London: Arnold, 184 p.

MYLES, F.; HOOPER, J.; MITCHELL, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, Oxford, UK and Cambridge, USA, v. 48, n. 3, pp. 323-363.

NEMSER, W. (1974). Approximative systems of foreign language learners. In: Richards, J. C. (Ed.). *Error analysis*: perspectives on second language acquisition. London: Longman., pp. 55-63.

NICKEL, G. (1998). The Role of interlanguage in foreign language teaching. *IRAL*, Berlin, v. 36, n. 1, pp. 1-10,.

ODLIN, T. (1993). Review of Selinker, 1992. *Language*, Washington, DC, v. 69, n. 2, pp. 379-383.

OXENDEN, C.; SELIGSON, P. (1996). *English file*: student's book 1. Oxford: Oxford University Press, 149 p.

PERRETT, G. (2000). Researching second and foreign language development. In: UNSWORTH, L. (Ed.). *Researching language in schools and communities*: functional linguistic perspective. London and Washington: Cassell, pp. 87-110.

PRAXEDES FILHO, P. H. L. (1996). Complexidade gramatical e densidade lexical: critérios marcadores da diferença entre narrativas orais e escritas. Fortaleza,. 277 p. Dissertação (Mestrado em Letras) – Centro de Humanidades, Universidade Estadual do Ceará.

SELINKER, L. (1974). Interlanguage. In: RICHARDS, J. C. (Ed.). *Error analysis*: perspectives on second language acquisition. London: Longman,. pp. 31-54.
_____. (1992). *Rediscovering interlanguage*. London and New York: Longman, 288p.

WASHBURN, G. N. (1994). Working in the ZPD: fossilized and nonfossilized nonnative speakers. In: LANTOLF, J. P.; APPEL, G. (Eds.). *Vygotskian approaches to second language research*. Norwood, NJ: Ablex,. pp. 69-81.

WIDDOWSON, H. (1979). *Explorations in Applied Linguistics*. Oxford: Oxford University Press, 273 p.

APPENDIX

KEY FOR DATA CATEGORIZATION:

Processes: Material– M; Mental– ML; Relational– R; Verbal– V; Existential– E

Regularly incomplete clauses (those categorized as idiosyncratically incomplete in SFG): non-finite clauses– NFC; anaphoric elliptical clauses– AEC; exophoric elliptical clauses– EEC; verb-less clauses– VLC.

Clauses without structure (those categorized by SFG as unanalyzable as for Transitivity, Mood, or Theme): minor clauses– MC.

Discourse markers: DM (these will be disconsidered since they are irrelevant to the aims of this study).

Incomplete clauses as for Transitivity or Mood: IC

Complete clauses as for Transitivity and Mood: CC

Simplification-related deviant complete clauses (simplification is deviation cause – criterion: from more complex to less complex): SDCC

Deviant complete clauses (simplification is not deviation cause): DCC

Language-acquisition-process type of incompleteness: LAP

Language-production-process type of incompleteness: LPP

Spoken narrative # 01,02,03... (rendered by Subject 1, 2, 3...): SN01, SN02, SN03, etc.

Written narrative # 01, 02, 03... (rendered by Subject 1, 2, 3...): WN01, WN02, WN03, etc.

Boundary indication for ranking clauses: ‖   ‖

Boundary indication for rank-shifted clauses: [[   ]]

Boundary indication for inserted clauses: ≪   ≫

DATA CATEGORIZATION OF 'SN07' AND 'WN07':

SN07 (SUBJECT 7)

CLAUSE 1: I driving the car in avenue (IC) (LAP)

	I	driving		the car	in + Φ + avenue
TRANSITIVITY	Actor	Process: Material		Goal	incomplete Circumstance
MOOD		Φ	drive		
	Subject	Finite lacking	Predicator	Complement	incomplete Adjunct
	incomplete Mood			incomplete Residue	

CLAUSE 2: and *bateu* another car (IC) (LAP)

	and	*(ele)* (Φ)	*bateu* (Φ)		another car
TRANSITIVITY	X	L2 Actor lacking	L2 Process: Material lacking		Goal
MOOD			past (Φ)	*bater* (Φ)	
	X	L2 Subject lacking	L2 Finite lacking	L2 Predicator lacking	Complement
	X	Mood lacking		incomplete Residue	

CLAUSE 3: I'm (IC) (LPP)

	I	'm + Φ		Φ
TRANSITIVITY	Actor	Process: Material lacking		Goal lacking
MOOD		present	Φ	
	Subject	Finite	Predicator lacking	Complement lacking
	Mood		Residue lacking	

144

CLAUSE 4: I'm broking in the leg in the in the my arm and my (IC) (LPP)

	I	'm broking		in the leg in the in the my arm and my + Φ
TRANSITIVITY	Actor	Process: Material		incomplete Goal
MOOD		present & past	break	
	Subject	erroneous Finite	Predicator	incomplete Complement
	erroneous Mood			incomplete Residue

CLAUSE 5: achieve my hand my head (IC) (LAP)

	Φ	achieve		my hand my head
TRANSITIVITY	Actor lacking	Process: Material		Goal
MOOD		present	achieve	
	Subject lacking	erroneous Finite	erroneous Predicator	Complement
	incomplete erroneous Mood		erroneous Residue	

CLAUSE 6: after I go to hospital (IC) (LAP)

	after + Φ	I	go		to + Φ + hospital
TRANSITIVITY	incomplete Circumstance	Actor	Process: Material		incomplete Circumstance
MOOD			present	go	
	incomplete Adjunct	Subject	erroneous Finite	Predicator	incomplete Adjunct
	incomplete Re...		erroneous Mood		...sidue

CLAUSE 7: but but is ok (IC) (LAP)

	but but	Φ	is		ok
TRANSITIVITY	X	Carrier lacking	Process: Relational		Attribute
MOOD			present	be	
	X	Subject lacking	erroneous Finite	Predicator	Complement
	X	incomplete erroneous Mood		Residue	

WN07 (SUBJECT 7)
CLAUSE 1: I driving the car (IC) (LAP)

	I	driving		the car
TRANSITIVITY	Actor	Process: Material		Goal
MOOD		Φ	drive	
	Subject	Finite lacking	Predicator	Complement
	incomplete Mood		Residue	

CLAUSE 2: and I bati another car. (IC) (LAP)

	and	I	bati (Φ)		another car.
TRANSITIVITY	X	Actor	L2 Process: Material lacking		Goal
MOOD			past (Φ)	bater (Φ)	
	X	Subject	L2 Finite lacking	L2 Predicator lacking	Complement
	X	incomplete Mood		incomplete Residue	

145

CLAUSE 3: I broker my leg and my arm. (CC)

	I	broker		my leg and my arm.
TRANSITIVITY	Actor	Process: Material		Goal
MOOD		past	break	
	Subject	Finite	Predicator	Complement
	Mood			Residue

CLAUSE 4: Afeter I go to hospital, (IC) (LAP)

	Afeter + Φ	I	go		to + Φ + hospital,
TRANSITIVITY	incomplete Circumstance	Actor	Process: Material		incomplete Circumstance
MOOD			present	go	
	incomplete Adjunct	Subject	erroneous Finite	Predicator	incomplete Adjunct
	incomplete Re...		erroneous Mood		...sidue

CLAUSE 5: but was all ok. (SDCC)

	but	was		all	ok.
TRANSITIVITY	X	Process: Relational		Carrier	Attribute
MOOD		past	be		
	X	Finite	Predicator	Subject	Complement
	X	simplified erroneous Mo...	Resi...	...od [L1 word order]	...due

146