3

The highly costly structures

As mentioned before, the difficulties exhibited by children who are linguistically impaired are diverse and all the subdomains of language can be affected. This thesis focuses specifically on deficits related to the syntactic domain, which is pervasively impaired and which may lead to other problems. The difficulties associated to this module are though subtle and may only be detected experimentally.

Psycholinguistic researches have indicated that some syntactic structures, such as passives, relatives and WH- and WH-+N interrogatives, involve considerable complexity, and some experts have been referring to them as highly costly structures.

This idea has been corroborated by the fact that such structures are acquired late in the typical language development, as well as because they are often impaired in cases of linguistic deficiencies. Being so, from a psycholinguistic theoretical perspective, it is possible to associate the late acquisition of those structures and the difficulties related to their processing, to their computational complexity, which may be analyzed by different theoretical approaches.

This work intends to investigate the complexity related specifically to relatives and to WH-+N interrogatives and, for that reason, it is important to analyze it according to different perspectives. The ideas supported by the Generative Linguistics, in the light of the Minimalist Program, and the Integrated Model of Online Computation (MINC) proposal are, therefore, analyzed below.

3.1

Linguistic framework

Generative Linguistics, developed from Noam Chomsky's ideas, supports that every human being has an innate knowledge of the properties that are common to the human languages. Chomsky develops a formal model and takes it

as a possible model of the intuitive language knowledge. He assumes that in order for the knowledge of a particular language to be acquired there must be an initial state (Universal Grammar) that restricts the set of hypotheses about the structure of language that might be formulated based on the input data, i.e. the set of combinatorial possibilities given a sequence of lexical items.

According to this perspective, the human being is provided with a faculty of language whose initial state restricts the possibilities of analysis of the linguistic input in the identification of the grammar of the language. Language is considered to be a specific domain to be developed from a given biological configuration and as a function of linguistic experience. The idea of a domain with specific resources is compatible with a modular conception of the mind.

The developments of the Generative Linguistics in the last four decades have led to the Minimalist Program. It corresponds to an extension of the Principles and Parameters theoretical framework.

The Minimalist Program intends to avoid redundancy, eliminating what is not really necessary to the model. The idea of *economy* is, therefore, really important in this perspective, so much so that this epistemological notion may be considered a type of principle, which has been guiding the development of the referred program.

From a theoretical perspective, the program intends to provide a concept of language reduced to its conceptually necessary properties, without which the object of study could not be the human language. As for methodology, it intends to simplify analysis, reducing the number of necessary operations and theoretical constructs within the grammar. The theory must, then, be parsimonious if the principles and operations are redundant.

In the light of such program, language is a perfect system that responds to external constraints optimally. Bearing in mind that the language faculty fits within the larger architecture of the mind (brain), it is the *interface levels*[1] (PF and LF – phonetic and logical form) between the internal language and the so-called performance systems, which determine the conditions that must be satisfied, in order for the result of the syntactic computation to be amenable to

[1] The interface levels are the only levels of linguistic description in the Minimalist Program. The term refers to the result of syntactic computation represented as a pair PF (phonetic form), LF (logical form) that is accessible to the systems that interact with language (sensory-motor and conceptual/intentional, respectively).

articulation/perception and to semantic interpretation/mapping in a referential context. Language interacts, then, with two performance systems: the articulatory-perceptual and the conceptual-intentional one. The first one accesses the phonetic representation of a linearized sequence of lexical items (*PF*), while the second one accesses the logical representation (of the content) (*LF*). Thus, to each generated expression, the language faculty must associate a level of representation (*PF* or *LF*) with each one of the performance systems.

> The sensorimotor systems, for example, have to be able to read the instructions having to do with sounds, that is the "phonetic representations" generated by the language. The articulatory and perceptual apparatus have specific design that enables them to interpret certain phonetic properties, not others. These systems thus impose legibility conditions on the generative processes of the faculty of language, which must provide expressions with the proper phonetic form. The same is true of conceptual and other systems that make use of the resources of the faculty of language: they have their intrinsic properties, which require that the expressions generated by the language have certain kinds of "semantic representations", not others. (Chomsky, 2000, p. 9)

A fundamental concept in Generative Linguistics is the Principle of Full Interpretation, according to which language must satisfy the impositions of the performance systems, since they must be able to interpret expressions of the language and make use of them for thought and action. The referred principle must be understood as the set of cognitive constraints imposed by the performance systems for the human language functioning. It establishes, therefore, that every linguistic representation must be concomitantly readable in both phonetic and logical interfaces.

According to this perspective, the architecture of language is constituted by a universal computational system and a lexicon. The first one recursively combines the items taken from the lexicon, generating complex expressions of sound and meaning (linguistic expressions such as phrases and sentences). The lexicon, in turn, is composed of elements which present semantic, phonological and formal (interpretable and uninterpretable) features.

It is relevant to mention that the formal features[2] play an important role in the constitution of natural languages and they are the only features the computational system is able to recognize.

[2] It is worth emphasizing that languages vary in relation to the properties of the formal features of the functional categories. So, in order to have access to the grammar of their community language and to initiate the language acquisition process, the child would have to identify such

In the Minimalist Program a derivational perspective is assumed. The derivation starts by the merging of two lexical items, selected from an array/ subarray of items from the lexicon (numeration) and proceeds by merging syntactic objects with lexical items selected from the numeration. The numeration contains all the lexical items to be used in the computation, each of them with an index indicating the number of times it is selected. The derivation stops when all the items have reached the index zero. The internal merging of a structure enables a constituent to move from one syntactic position to another. The movement operation (or internal merge) has been presented as a source of computational cost (Chomsky, 1995).

Once the derivation ends, the information concerning the correspondence between structural positions and linear ordering is spelled out to PF and LF.

3.1.2

Computational cost in the Minimalist Program

Move[3], also called *Internal Merge*, is a fundamental operation for understanding a large number of syntactic phenomena of the natural languages. The operation in question displaces constituents in the course of a derivation, leading to long distance syntactic relations whereas abiding by locality constraints.

There are three types of movement, according to the Minimalist Program: Head movement, the movement of an element from a head position to another head position; A-movement (or Argument movement) and A'-movement. In the case of A-movement, the element is moved to an argument position as in passive sentences (the logical object moves from the position in which its theme/patient

features. Formal features play, therefore, an important role in the language acquisition process, since they are visible in the phonic interface, in what is regular, and encode distinctions considered as grammatically relevant.

[3] The operation Move is applied as follows: at first, the Computational System produces a copy of the constituent that will be moved; then, the Computational System moves a copy of the constituent to the target position, where it merges with another syntactic element; and, finally, the copy left on the base position is deleted. It is important to note that this operation is triggered as a consequence of the following fact: the presence of an uninterpretable feature, which is the probe, triggers the search for this element. Associated with this uninterpretable feature, there is the EPP feature, which is responsible for bringing this element into the domain where the probe is located.

role is assigned to occupy the subject position). That is to say that the constituent is displaced to the position of a purely syntactic argument. As for A' movement, the constituent is moved to a non-argument position, that is, a position where no argument is selected. This movement is typical of interrogatives, relatives and topicalizations. As stated by the Minimalist Program, *Move* is a costly operation and, among the types of movement discussed above, A movement and A' movement are the most costly ones, since both of them involve the movement of a phrase to a specifier position (specifier of TP; of CP). It is also important observing that the types of movement in question are, though, equally costly, according to this perspective.

In other words, as mentioned above, relatives, WH, and WH+N questions involve the A' movement that, as claimed by the Generative Linguistics, may be considered a costly operation. Thus, the derivation of a WH+N question, such as "Which book did the teacher buy?", may be observed in the following figure.

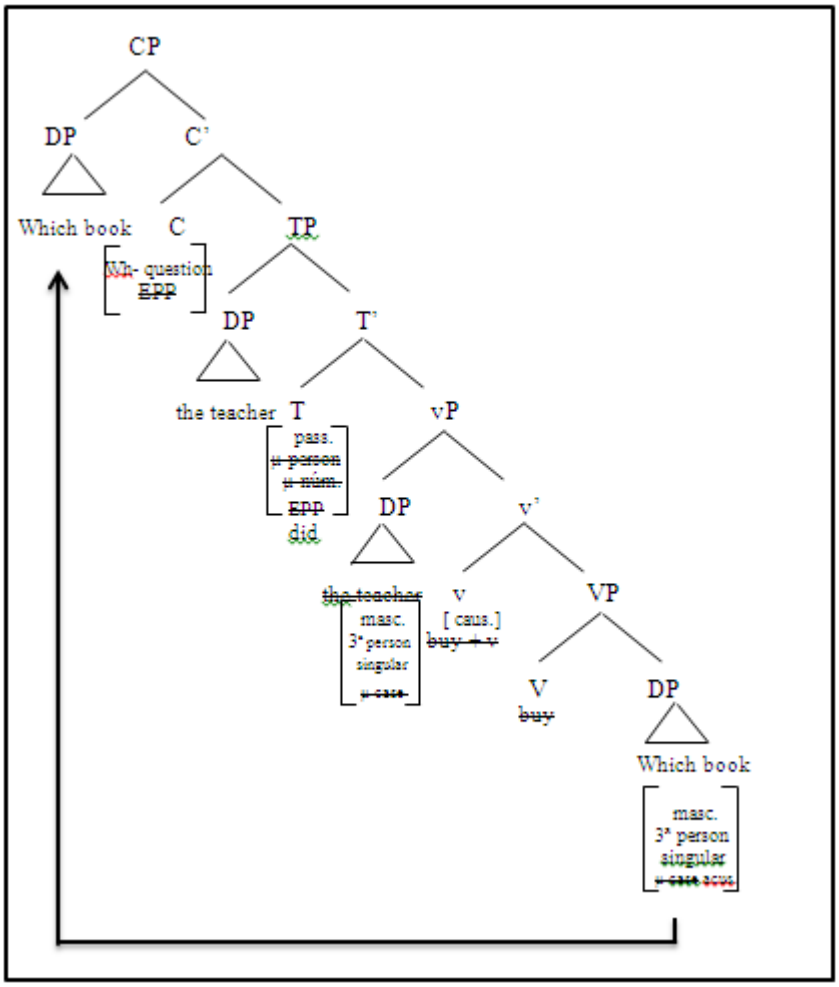

Figure 1: WH+N question derivation.

Figure 2, in turn, illustrates the derivation of the relative structure "The dog that the boy chose was brown."

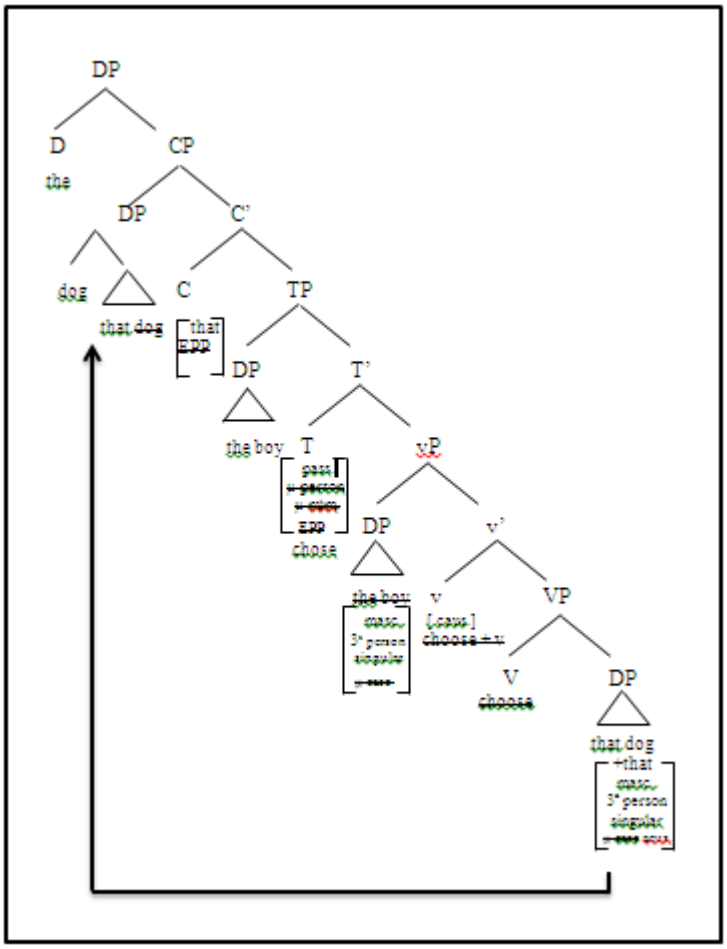

Figure 2: Part of a relative clause derivation.

3.2

The Integrated Model of Online Computation (MINC)

The Integrated Model of Online Computation (MINC) (Corrêa & Augusto, 2007; 2011) intends to provide a procedural (psycholinguistic) model of online (real time, incremental) computation, i.e. the syntactic computation carried out during sentence production or comprehension, in the light of a formal model of syntactic computation as characterized in a minimalist derivation.

> The integrated model of on-line computation (MINC) incorporates much of the operations of the language computational system characterized in the MP (namely, Merge, Agree, Move (Merge+Copy)) as necessary operations for the syntactic formulation of sentences in production, and for sentence parsing in comprehension. (Corrêa, Augusto & Foster, 2012, page 10).

According to MINC, the operations triggered by a universal computational system, as proposed by the Minimalist Program, are incorporated and/or adapted in order to characterize algorithmic procedures, which would be implemented in production and comprehension (Corrêa & Augusto, 2009, page 52).

In this perspective, a basic scheme of linguistic production would involve: the retrieval of lexical elements from a speech intention, a "message"; the linguistic formulation, that is, the positioning of lexical elements in syntactic structures; morphophonological coding; and articulatory planning.

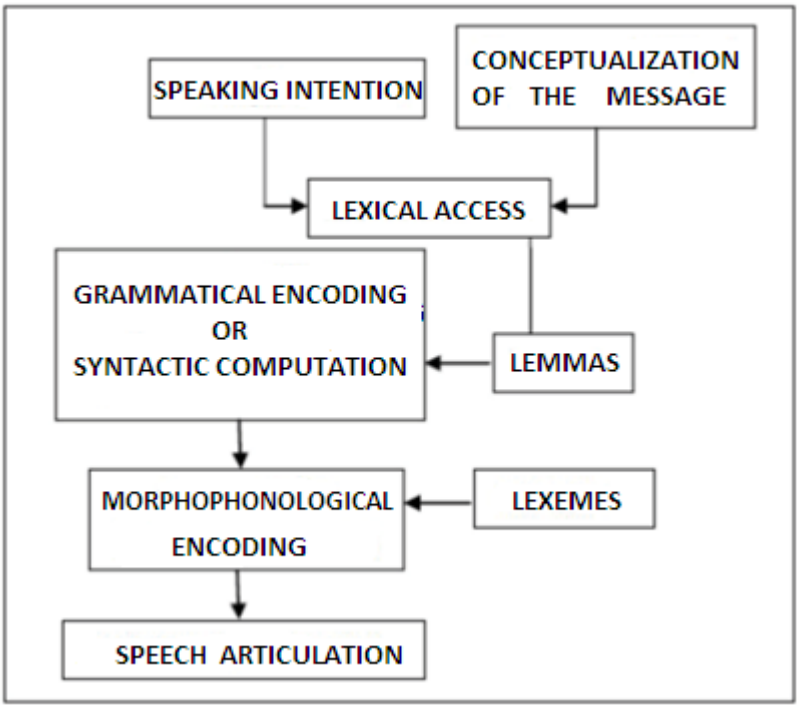

Figure 3: A basic scheme of linguistic production (Corrêa & Augusto, 2013, p. 39).

Conversely, the comprehension process would involve an inverse scheme, which would include: the acoustic signal perception and its segmentation; lexical recognition; syntactic analysis, that is, the positioning of the recognized elements in a hierarchical structure; and semantic interpretation, which would be integrated with the listener's other knowledge bases.

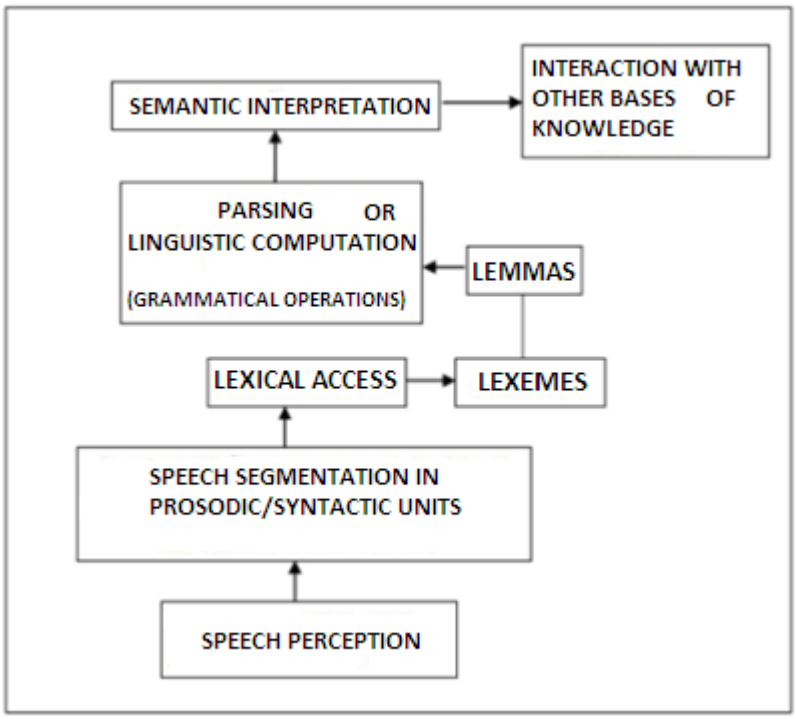

Figure 4: A basic scheme of linguistic comprehension (Corrêa & Augusto, 2013, p. 40).

As previously noted, according to the minimalist model, a syntactic derivation results in two levels of representation: Phonetic Form (*PF*) and Logical Form (*LF*). The former interfaces with the articulatory-perceptual system, while the latter interfaces with the conceptual-intentional system. Based on this conception, MINC assumes that all the necessary information for the parsing and for the semantic interpretation of a linguistic utterance becomes visible in the interfaces of the language with the systems that act in the linguistic processing (Corrêa & Augusto, 2013, page 40).

In the Minimalist Program, the formal features of the lexical elements trigger the operations of the Computational System. These features are associated with both lexical categories and functional categories of the lexicon. As noted by Corrêa & Augusto (2013, p. 42), the formal features associated with functional categories define the type of reference to be made to entities (in D - determinants), to events (T - time and Asp - aspect), as well as the type of illocutionary force (C - complementizador) of the statement.

Unlike in a formal (minimalist) derivation, the syntactic computation is as a bidirectional process in MINC. Thus, it is believed that the access to elements of

functional categories would trigger the construction of basic structural skeletons in a *top-down* form; whereas the elements of lexical categories would generate structures through a *bottom-up* derivation.

Considering that the model in question is aimed at making explicit the operations present in the parsing and also in the formulation of linguistic utterances in real time, it is important to consider the different types of syntactic movement that can be generated by the application of the *Move* operation.

According to MINC, the syntactic movements that characterize the positioning of elements in the canonical order of a language would lead to the formation of **simultaneous copies**, which would not imply computational cost. On the other hand, the movement of constituents generated by discursive demands would entail the formation of **sequential copies**, which, in turn, would imply computational cost.

Even though MINC predicts cost for operations generated by discursive demands, it is not clear to what extent a behavioral manifestation of processing cost is necessarily obtained. It is known that object relative clauses and object interrogatives (particularly WH+N questions) are more costly than their subject counterparts. The presence of the WH+ move feature may not be the most crucial determinant of computational cost. Actually, cost may be entailed by the need to hold the head noun (with this feature) active in working memory while the processing of syntactic relations is carried out. The processing of a new sentence would be expected to bring more demand if the subject is filled in lexically, as in object structures. The predicted cost derived from this presence of the subject as an intervening element between the moved constituent and its original position is compatible with the intervention hypothesis, formulated in terms of universal principles, which will be discussed below.

So, in this perspective, it becomes possible to predict that both object relatives and object WH+N questions, for involving the last type of movement and for including a possible intervening element, give rise to increasing computational cost, which could lead to difficulties in language acquisition, take longer in deviant development and even be insurmountable in pathological cases.

3.2.1

Interrogatives

As stated by MINC, the syntactic computation conducted during the parsing of a WH+N question, such as "Which book did the teacher buy?", is illustrated in the figure below.

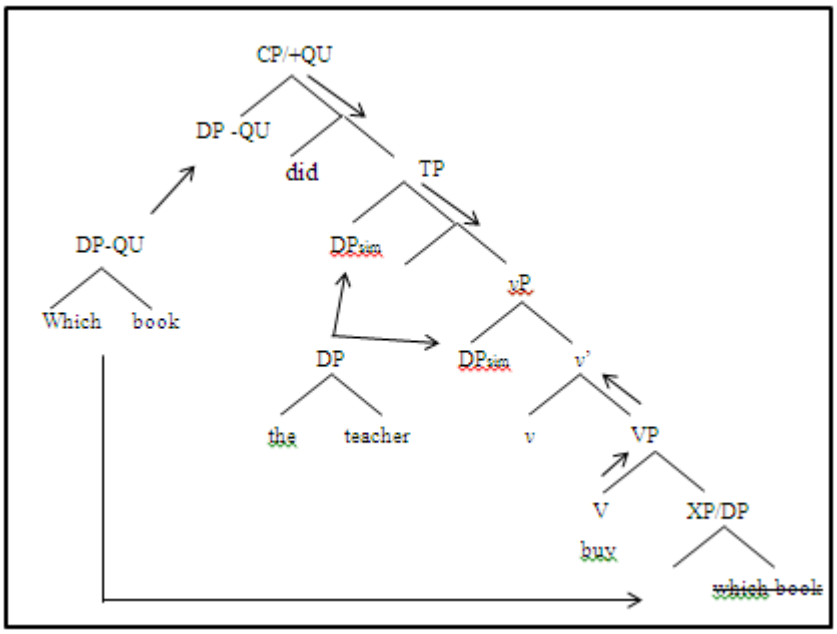

Figure 5: Syntactic computation of a WH+N question.

In this example, the first constituent to be processed from left-to-right is the phrase *Which book*. Thus, an interrogative DP is generated, which allows the *top-down* generation of an interrogative CP and a TP, which predicts a sentence with specified grammatical tense.

In such case, the WH element must be maintained active in the working memory until the first empty syntactic position, in which its thematic role can be assigned, is identified. Holding this element active implies that sequential copies will be generated while the structure is computed. This computation involves the generation of a subject DP by means of simultaneous copying. The presence of elements with similar features generated and maintained in parallel derivational spaces, until they occupy their hierarchical positions, would make the process costly (Corrêa &Augusto, 2013, p. 49).

3.2.2

Relatives

According to MINC, relative clauses also demand high computational cost. In the following figure, it is possible to observe part of the syntactic computation performed during the parsing of the sentence "The dog that the boy chose was brown".

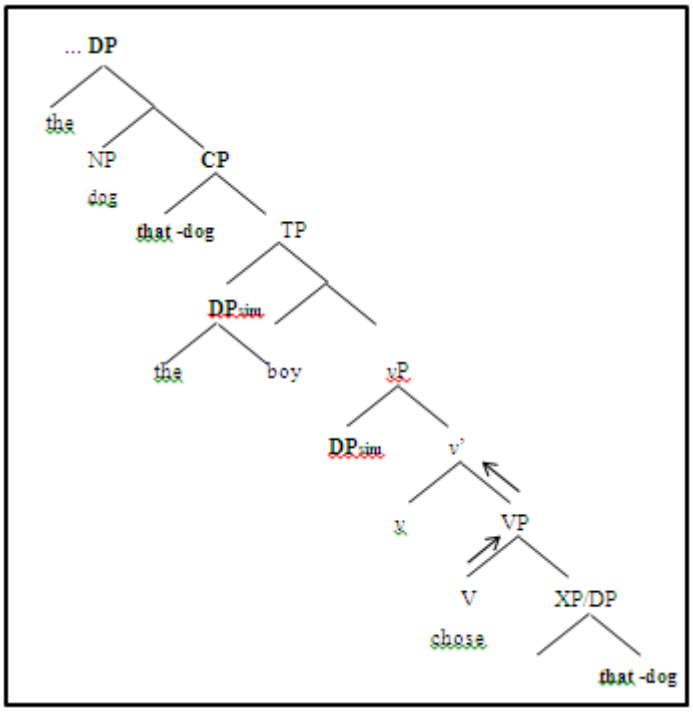

Figure 6: Syntactic computation of a relative sentence.

In the previous example, the relative pronoun "that" would indicate the need to keep the recent analyzed DP active in the working memory, so that it would be integrated to the following structure.

As explained by Corrêa & Augusto (2013, p. 52), in this case, the *that* signals the need to generate a top-down CP / TP to which the structure projected by the relative verb is coupled. The DP held in memory can then be retrieved to fill an empty position (subject, object, etc.) in that structure.

At first, *that-dog* could be processed as the subject of the relative sentence. However, the DP "the boy" invalidates this interpretation and a new DP is generated. So, the *that-dog* copy must be maintained active in the working

memory until its syntactic and semantic role is identified. This process would, therefore, imply high processing cost. That is, this intervening DP imposes an additional demand since it needs to be generated at the same time that the DP *that-dog* needs to be kept in a "memory box" until its position is detected (Corrêa & Augusto, 2013, page 52).

According to the proposal at stake, both object WH+N interrogatives and relatives overload the working memory, since **sequential copies** are triggered by the presence of a feature such as [+move] in the DP headed by WH, which is structurally similar to the intervening element. In production, this feature promotes the fronting of the WH-phrase. In comprehension, the identification of the WH word at the beginning of the clause is taken as a cue implying that this element must be active in working memory until the position where its thematic role can be recognized is identified.

As mentioned above, psycholinguistic researches have indicated that the so-called highly costly structures, such as the ones analyzed in the present work, may imply difficulties for both typical and deviant linguistic development. It is necessary, therefore, to reflect on these structures, in order to understand what factors may be related to their complexity.

Formal models of language are not aimed to account for the production and comprehension processes. MINC, however, incorporates the theoretical assumptions of the Minimalist Program, while integrating them into psycholinguistic studies of language processing. In addition, such model is an attempt to make explicit how syntactic computation is developed in real-time language production and comprehension processes.

The referred approach can be quite productive, considering that from the characterization of real-time syntactic computation, it becomes possible to understand the reason why the structures in question are so costly. Thus, potential ways of remedying the difficulties related to the processing of these structures may be proposed, such as the planning of language stimulation activities for children with syntactic difficulties.

3.3

The intervention hypothesis

Syntactic relationships are generally established locally, that is, the elements of a relation of dependency must be located in the same domain of the syntactic structure. Locality constraints are intended to prevent the generation of structures, by the computational system, which cannot be processed in the interfaces of the language with the performance systems. By observing the following object relative clause, however, it is possible to notice that, in such case, the presence of an intervening element may compromise a local relation, if the crucial feature that distinguishes "the boy" from "the teacher" (namely, the +WH feature) is not taken into account (either due to an impairment preventing its access/recognition or to relatively great number of features shared by the object and the subject of the relative clause)[4].

Example: The boy that *the teacher* called [the boy] went to the bathroom.

In a structure like the one above, the DP "the teacher" acts as an intervening element between the moved element "the boy" and its original position. In the comprehension of such a sentence, the empty position (the position of a copy of the moved element) must be filled in by the moved element retrieved as a gap is recognized. The intervening element is nevertheless also active in working memory and may be recovered in such position thereby generating an incorrect interpretation of the sentence.

As observed above, in MINC, computational cost was associated to the movement of constituents generated by discursive demands, which cause changes in the canonical order, as well as a need to keep an active element in working memory until a copy of the moved element can be identified. Some studies (Gordon, Hendrick & Johnson, 2004; Warren & Gibson, 2002; Friedmann,

[4] It refers to the principle of Relativized Minimality: a local relation between X and Y cannot be established in the configuration X ... Z ... Y ... (where X c-commands Z and Z c-commands Y), if there is intervention from Z, and if Z shares with Y all the specifications of the formal features relevant to the syntactic operation at stake. Otherwise there is no impediment (see Rizzi, 1990, 2013).

Belletti & Rizzi, 2009; Ribeiro, 2012) suggest that the processing cost of this type of structure may be minimized the more dissimilar the features of the intervening element and the moved element are, e.g. if the former is a full DP and the latter is a pronoun.

According to Warren & Gibson (2002) and Gordon, Hendrick & Johnson (2004), the maintenance of two (or more) referentially similar elements in working memory would result in higher processing costs. Warren and Gibson (2002) and Ribeiro (2012) suggest that pronouns unlike full DPs and proper names in subject position would facilitate processing, since they would have less complex structures.

As for young children, SLI children and agrammatic adults, it has been suggested that the Principle of Relativized Minimality (see note 5) is overextended (Friedmann, Belletti & Rizzi, 2009) or generalized (Grillo, 2008) thereby hindering the comprehension of the sentence. It should be noticed that either from the perspective of sentence processing (when reference takes place) or from the perspective of an overextension of locality principles (constraining the form of possible grammars), object WH+N and RCs are predicted to be costly.

From the point of view of the psycholinguistic literature there is the interference of the subject of the structure, in terms of the introduction of a new referent, as well as of the complexity of its form. From the point of view of the linguistic intervention hypothesis, it refers to a principle that supposes the blocking of syntactic processes when there is a total sharing of features between the intervening element and the element moved. In the case of these structures, there would be no blockage, but possibility of high cost, since the crucial distinction would be in the + WH feature.