Artificial Intelligence in Medicine 61 (2014) 89–96

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

A methodology for the characterization and diagnosis of cognitive impairments—Application to specific language impairment

Jesús Oliva*, J. Ignacio Serrano, M. Dolores del Castillo, Ángel Iglesias

Bioengineering Group, Spanish National Research Council (CSIC), Carretera de Campo Real, km. 0,200, La Poveda, Arganda del Rey, CP: 28500 Madrid, Spain

ARTICLE INFO

Article history:
Received 15 February 2013
Received in revised form 4 March 2014
Accepted 3 April 2014

Keywords:
Computational cognitive modeling
Machine learning
Specific language impairment
Automatic diagnosis

ABSTRACT

Objectives: The diagnosis of mental disorders is in most cases very difficult because of the high heterogeneity and overlap between associated cognitive impairments. Furthermore, early and individualized diagnosis is crucial. In this paper, we propose a methodology to support the individualized characterization and diagnosis of cognitive impairments. The methodology can also be used as a test platform for existing theories on the causes of the impairments. We use computational cognitive modeling to gather information on the cognitive mechanisms underlying normal and impaired behavior. We then use this information to feed machine-learning algorithms to individually characterize the impairment and to differentiate between normal and impaired behavior. We apply the methodology to the particular case of specific language impairment (SLI) in Spanish-speaking children.

Methods and materials: The proposed methodology begins by defining a task in which normal and individuals with impairment present behavioral differences. Next we build a computational cognitive model of that task and individualize it: we build a cognitive model for each participant and optimize its parameter values to fit the behavior of each participant. Finally, we use the optimized parameter values to feed different machine learning algorithms. The methodology was applied to an existing database of 48 Spanish-speaking children (24 normal and 24 SLI children) using clustering techniques for the characterization, and different classifier techniques for the diagnosis.

Results: The characterization results show three well-differentiated groups that can be associated with the three main theories on SLI. Using a leave-one-subject-out testing methodology, all the classifiers except the DT produced sensitivity, specificity and area under curve values above 90%, reaching 100% in some cases.

Conclusions: The results show that our methodology is able to find relevant information on the underlying cognitive mechanisms and to use it appropriately to provide better diagnosis than existing techniques. It is also worth noting that the individualized characterization obtained using our methodology could be extremely helpful in designing individualized therapies. Moreover, the proposed methodology could be easily extended to other languages and even to other cognitive impairments not necessarily related to language.

1. Introduction

The characterization and diagnosis of cognitive impairments are, in most cases, problematic. There are two main reasons for their difficulties: heterogeneity and overlap. Many cognitive impairments present a general behavioral profile inside which we can find many different behavioral trends for each patient. These individual differences are crucial for the design of personalized therapies and could provide important clues to better understand each deficit. Clinical evidence shows the effectiveness of early and individualized intervention for very different cognitive impairments and, in particular, for many communication disorders [1]. However, there are very few studies that consider these individual differences. The other main problem is the high overlap presented by many cognitive impairments. Many disorders co-occur and share similar symptoms and some aspects of their behavioral profiles.

All these factors motivate the design and development of new methodologies for the characterization and diagnosis of cognitive impairments that consider individual differences. Existing techniques just focus on behavioral variables. However, as stated above, behavior is in most cases heterogeneous and is shared by many

* Corresponding author. Tel.: +3491 3363061.
E-mail address: jesus.oliva@csic.es (J. Oliva).

http://dx.doi.org/10.1016/j.artmed.2014.04.002
0933-3657/

cognitive impairments. Therefore, focusing solely on behavioral variables might not be the most appropriate way to design such methodologies. It might be more appropriate to investigate the cognitive mechanisms underlying normal and impaired behavior. Finding patterns at the cognitive level might be very useful for the characterization and diagnosis of the impairment. The tool we propose to investigate the underlying cognitive mechanisms is computational cognitive modeling.

In this paper, we propose a methodology for the characterization and diagnosis of cognitive impairments and apply it to the particular case of specific language impairment (SLI) in Spanish-speaking children. The approach is based on building a computational cognitive model of a task in which children with impairment show differences with normal children. Then the different parameters of the model are used to train machine-learning algorithms to support the processes of diagnosis and characterization. We applied the methodology to an existing database of Spanish-speaking children. The results show that the information obtained from the computational cognitive models is highly informative for characterization and diagnosis. Additionally, the results show the utility of this methodology to support or reject existing theories on the causes of the impairment. It is also worth mentioning that this methodology could easily be extended to other languages and even to other cognitive impairments not necessarily related to language.

2. Background

SLI is generally defined as a developmental disorder of language ability in the absence of factors that usually affect language learning, such as hearing impairments, low non-verbal intelligence or neurological damage [2]. Numerous studies have investigated the cognitive profile of children with SLI and have attempted to characterize the impairment and establish a standardized diagnosis mechanism. However, none of these objectives have been achieved to date. In the following subsections we briefly review the existing theories on the disorder and the common diagnostic techniques.

2.1. Existent theories on SLI

SLI has been extensively studied from many different perspectives. However, there is no unified account for its particular developmental profile. There are two main reasons for this lack: heterogeneity and overlap. SLI presents a broadly heterogeneous profile [2]. Different individuals tend to show wide differences in the severity of the disorder, the areas of language affected by the disorder and how these areas are affected within the same child over time [3]. This variation has led to the definition of different SLI subgroups and even different definitions for each individual profile within SLI [4,5]. The other main reason is the high overlap of SLI with other cognitive impairments. SLI co-occurs with certain other disorders such as semantic-pragmatic disorder and the autistic spectrum [4], attention deficit hyperactivity disorder [6], and some motor-related disorders [7]. Furthermore, contrary the first theories that tried to account for SLI, it has now been proved that this impairment is not just restricted to language. Many other cognitive functions are impaired such as working memory [8] and some motor skills [7]. This high heterogeneity and overlap make it very difficult to differentiate SLI from other cognitive impairments and to distinguish between different subcategories grouped under the SLI profile. Therefore, there is no unified account for the particular developmental profile of SLI. There exist three main theories that attempt to account for this developmental profile: the grammar specific deficit (GSD), the processing deficit (PD) and the procedural deficit hypothesis (PDH).

2.1.1. Grammar specific deficit

GSD theories consider SLI to be due to a deficit in the processing of grammar. The most famous example of this theory is most likely the extended optional infinitive (EOI) hypothesis [9]. Typically developing children go through a stage in which they omit verbal suffixes such as tense, person and number agreement markers [10]. The EOI hypothesis argues that this stage of immature grammar is much longer in SLI children. Other authors claim that the problems with grammar are due to the under-representation of grammatical and linguistic features [11,5]. Although these theories account quite well for the language profile of SLI children, they cannot explain all the linguistic phenomena detected either in English or in other languages. Moreover, any purely grammatical explanation cannot account for the problems presented by SLI children when performing non-linguistic tasks.

2.1.2. Processing deficit

Some authors attribute the problems associated with SLI to a processing deficit not related to language. This particular processing deficit could be general or specific to very concrete systems such as the phonological system (see [12] for a brief review). General processing deficit theories attribute the problems of SLI children to a somewhat limited processing capacity [13,14]. The generality of the impairment proposed by these theories explains many of the deficits of the SLI profile. However, they fail to account for some linguistic and non-linguistic impairments [2,15]. There are also some processing deficit theories that consider that the SLI profile to be caused by a specific deficit localized in the phonological system, affecting the working memory [16]. Again, these hypotheses fail to account for all the impaired linguistic and non-linguistic functions found in SLI [2,15].

2.1.3. Procedural deficit hypothesis

According to the declarative-procedural model [17], there exists a dissociation in the cognitive system between declarative and procedural memory. Declarative memory stores semantic knowledge, while procedural memory handles learning, grammar and rule-based skills. The PDH [18] is an extension of this theory that proposes that SLI is basically a disorder of the procedural system. Declarative memory tries to compensate for this deficit, but the result is still an impaired developmental profile [15]. Thus, this model explains the linguistic problems related to the acquisition of syntax and also the motor deficits found in SLI, (by maintaining that those deficits are related to implicit sequence learning tasks that rely on the procedural system). The PDH seems to be the most complete account of SLI to date. However, some authors note that not all of its conclusions are necessarily true [19] and that a statistical learning account of SLI could better explain the behavioral data [20].

All the existing theories capture some features of the behavioral profile of SLI. However, there is no hypothesis that accounts for the wide range of linguistic and non-linguistic problems found in SLI and for the individual differences present in the SLI profile. This fact could suggest the existence of different subtypes of SLI and the consequent need for individualized diagnostic techniques.

2.2. Current diagnostic techniques

SLI diagnosis is usually based on both inclusionary and exclusionary criteria. The general approach for assessing language development uses language tests such as inflection tasks, vocabulary tests or non-word repetition tasks, among many others [2]. Typically, children performing significantly lower than average on two or more of these measures are diagnosed as impaired. Some other observable variables are taken from transcriptions of children's spontaneous speech, such as the mean length of utterance

or percentage of different errors. Finally, there are measures based on non-linguistic tasks such as hearing sensitivity, interaction with people and objects or neurological status [2]. Some of these measures are used for each assessment test. However, some recent work has reported out that these assessment tools might not be completely appropriate [21].

Additionally, artificial intelligence, machine learning and statistical techniques have been used for the neuropsychological assessment and diagnosis of different impairments. The idea underlying these approaches is to extract different rates and scores from narratives or psycholinguistic tests and then use those features to build classifiers using different machine learning techniques. For example, some authors propose the use of alignment techniques to automatically extract narrative recall scores for the diagnosis of mild cognitive impairment [22,23]. The classification results from using a support vector machine are comparable to the results obtained with the manually extracted features. Promising results have also been reported for the detection of pragmatic language impairment in children with autism using a similar approach [23]. Moreover, the diagnosis of SLI has also been studied using information extracted from narratives and from different natural language processing and machine-learning techniques such as fuzzy cognitive maps [24], language models [25] or standard classifiers [26,27] with promising results. All these works show the potential utility of natural language processing and machine-learning techniques in this area. Nevertheless, these methods still need improvement for use as a standard diagnosis method.

Common to all of these approaches is the fact that they work with behavioral observations but, as stated before, SLI behavioral patterns are highly heterogeneous and are also found in other disorders. Given these conditions, behavior does not seem to be sufficient. Our contention is that access to the mechanisms underlying normal and impaired behavior could facilitate the diagnosis process. Therefore we propose that computational cognitive modeling could be an extremely useful tool to gain access to those mechanisms.

3. Methodology for the characterization and diagnosis of cognitive impairments

As noted in the previous section, two key factors, overlap and heterogeneity, make it difficult to diagnose cognitive impairments. To avoid these two problems, any methodology for the diagnosis of cognitive impairments should have two salient features: generality and individualization. The methodology should be appropriate to diagnose several different cognitive impairments and, at the same time, it should consider the individual differences that are usually present in those impairments. Computational cognitive models of cognitive impairments usually attempt to mimic group performance, excluding the particular behaviors of each individual. However, individualized characterization and diagnosis is extremely important for the design and development of more effective therapies [28]. Here we present a methodology that achieves these two objectives (generality and individualization) and apply it to the particular case of SLI. The different stages of the methodology are explained below.

- *Define a target task*: In this stage, we find a task or set of tasks in which the patients show difficulties and behavioral differences with respect to healthy individuals.
- *Computational cognitive modeling*: The next step is to build a computational cognitive model for the target task. The psychological plausibility of the model is a key point. The model should be able to show normal and impaired behavior, but how the model produces those behaviors is also of great importance because that information will be used in the diagnosis process. The better the model mimics human behavior, the more useful the information obtained will be.
- *Subject modeling*: Given the greatly heterogeneous profile of many cognitive impairments, any diagnostic methodology should consider individual differences. Our proposal is to obtain, for each individual, the combination of parameter values of the computational cognitive model that best matches the behavior of that individual. Therefore, this stage of the methodology uses an optimization algorithm for the parameter values.
- *Application of machine learning techniques*: The final stage of the methodology consists of applying different machine learning techniques using the information derived from the computational cognitive model. This stage tries to demonstrate the importance of the two key elements of the methodology: the information derived from the computational cognitive model and the techniques used to process this information. We want to show that the cognitive variables have more predictive power than the observable variables and that the machine learning techniques can yield better results in terms of classification and characterization than the classical approaches.

4. Application to SLI

As stated in Section 2, SLI is a good example of a cognitive impairment that is difficult to characterize and diagnose. In this section we explain how we applied the methodology proposed in the previous section to the case of SLI.

4.1. Target task

SLI children have problems with many grammatical tasks, especially verbal morphology [2]. Those problems have been extensively studied in English and have also been reported in many other languages. Although such errors are less frequent in Spanish [29,30], they are still a good marker for SLI [31]. Thus we decided to choose a verb inflection task. Given the infinitive form of many different verbs, the model has to inflect them with different combinations of mood, tense, aspect, number and person and acquire inflections through development.

In this paper, we will use the study by Simon-Cereijido and Gutiérrez-Clellen [32] on Spanish-speaking SLI children. They compiled a database of 48 Spanish-speaking children, 24 with typical language development (TLD) and 24 with language impairment (LI). The two groups were matched in age (TLD 4;6, LI 4;5) and socioeconomic status. The children in the LI group were selected based on three criteria: evidence of parent or teacher concern, clinical judgment by speech-language pathologists and low scores on morphosyntax tests. None of the children had any other impairment such as mental retardation, hearing impairment or motor difficulties. The children were asked to tell a story based on two different wordless picture books: *Frog goes to dinner* [33] and *Frog on his own* [34]. The samples were transcribed using SALT [35] and analyzed to compute different grammatical measures related to verbal morphology (rates of correct use, omissions, over-regularizations and substitutions of number, person and tense), articles (rates of correct use, omissions and substitutions of number or gender), clitics (rates of correct use, omissions and substitutions of case, person gender or number) and argument structure (rate of use of transitive structures, rate of use of ditransitive structures and theme argument omission rate). Table 1 shows the scores obtained by each group on these grammatical measures. Complete details on the participants and procedure can be found in Simon-Cereijido and Gutiérrez-Clellen [32].

Table 1
Language performance for the two groups of children involved in the study of [32]. SD stands for standard deviation.

Measure	Typical development		Language impairment	
	Mean	SD	Mean	SD
Mean length of utterances	5.68	1.01	4.44*	1.54
Ungrammatical sentences	0.17	0.04	0.35*	0.11
Correct use of articles	0.96	0.06	0.79*	0.22
Correct use of verbs	0.97	0.02	0.90*	0.08
Correct use of clitics	0.92	0.05	0.80*	0.17
Correct use of theme arguments	0.95	0.04	0.90	0.10
Proportion of ditransitive structures	0.08	0.03	0.05	0.04

* Significant difference ($p < 0.01$) with the corresponding value on healthy children using a two tailed t-test.

4.2. Computational cognitive model

We built a computational cognitive model using the cognitive architecture ACT-R [36]. A detailed description of a previous version of the model can be found in [37]. We describe its key features below.

Representations. The model uses semantic, morphological and phonological information. The semantic information is represented using a local encoding scheme with a unique code for each verb root. The phonological information uses the distributed phonological encoding scheme from [38]. The morphological information comprises of grammatical features such as conjugation, number, person, mood, tense or aspect. All this information is organized in *memory chunks*, the main units of declarative knowledge in ACT-R. The memory chunks consist of a list of *information type - value* pairs. In our model, each verb form is represented by a chunk that stores the semantic, morphological and phonological information mentioned above.

Mechanisms. The model is based on two general strategies: memory retrieval and analogy. Using these two initial strategies, the model can acquire the regular rules and the irregular exceptions using only the examples from the input vocabulary. The general mechanisms of ACT-R guide these two initial strategies through development. Of particular interest are the production compilation and the partial matching mechanisms. Production compilation allows the model to acquire new rules (for example, the regular rules that govern inflectional morphology) solely by inducing general rules based on observed cases. Partial matching controls the retrieval process. The item retrieved for each query need not exactly match the retrieval query slots. Therefore, partial matching controls how sensitive retrieval is to each feature.

Parameters. The mechanisms of the model are controlled by a series of parameters that shape its behavior. These parameters form four main groups:

- Declarative memory: *RT* is the retrieval threshold. It controls which chunks can be retrieved from memory and which ones cannot be remembered. *ANS* defines the noise introduced into the memory retrieval process. *BLL* controls how fast memory elements are forgotten.
- Procedural memory: The parameter α controls the learning rate of new rules and the reinforcement of existing ones. *EGS* controls the noise in the process of selecting a rule to execute.
- Grammatical processing: The parameter γ_m controls the noise introduced into the perception of morphological features. The parameter γ_m-decay controls how the morphological noise decays in the course of development. *Partial matching parameters* (C-PM, NP-PM and MTA-PM) control the sensitivity of the model to each grammatical feature (conjugation, number-person and mood-time-aspect, respectively) when retrieving a verb form from memory.
- Phonological processing: The parameter γ_{ph} controls the noise introduced into the perception of phonological features. The parameter γ_{ph}-decay controls how the previous parameter decays in the course of development and *Ph-PM* controls the sensitivity of the model to phonological features when retrieving a verb form from memory.

Input vocabulary: We used the Spanish verb inventory [39], which contains frequencies for the present and past tense forms of 50 of the earliest acquired Spanish verbs. The input for the model consists of 220 immediate past forms and 248 present tense forms of the Spanish verb inventory. Each of these forms has an associated frequency of use in children's language. Verb forms were presented to the model following this frequency distribution. Following the design of Taatgen and Anderson [40], every 2000 simulated seconds, the model 'hears' two words (which means that two chunks representing correct verb forms are stored in declarative memory) and is asked to 'produce' one verb form. The method and vocabulary used to measure the performance of verb inflection in our model is not the same as the ones used for the children. However, both of them try to measure the ability to inflect known verbs when they are produced spontaneously. The vocabulary we used is based on estimations of the earliest acquired verbs. It should therefore be a good sample of children's standard vocabulary [39]. The method simulates spontaneous production rates in children, as it is based on real frequency distribution estimations. Thus it is suitable to measure those rates, just as in Simon-Cereijido's study [32].

4.3. Subject modeling

As stated above, it is very important to obtain an individualized diagnosis for each patient. Therefore, we do not focus on group performance but on modeling each subject individually. We used an evolutionary strategy [41] to obtain the combination of parameter values of the computational cognitive model that best fit the behavior of each of the 48 individuals in the database.

The genotype used consists of the 13 parameters of the ACT-R model mentioned above. To constrain the search space to psychologically plausible values we used the database proposed by [42]. In this work, the authors proposed a database that collects the parameter values of a representative range of published ACT-R models. Therefore, our model parameters are set by comparing them with many other models of very different tasks. However, because we are also trying to model individuals with impairment, the range allowed for each of the parameters to be defined as the average value ± five standard deviations. Thus, we constrain the search space to psychologically plausible values while allowing the simulation of impaired behavior. The value of five standard deviations was chosen to reflect theories that highlight that impaired behavior is the tail end of a normal distribution of cognitive capacities [38]. Our future work includes the optimization of this value. The fitness function used was the minimum mean square error between the children's correct production rate vector and the model's correct production rate vector at the same simulated age. The rest of the elements of the evolutionary strategy are well known:

- *Mutation*: We decided to use a non-correlated mutation with n step sizes. Therefore, in addition to the 13 genes, we also used 13 endogenous parameters to control mutation. The selection of the mutation strategy is not easy but some authors suggest starting

from this strategy and trying other options only if the results are not good enough in terms of fitness or efficiency [43].

- *Crossover*: We used, as is typical with evolutionary strategies [41], a discrete crossover for the genes and an intermediate crossover for the endogenous parameters. In both cases, we used two parents to generate each individual.
- *Selection of individuals*: We used a (μ, λ) selection strategy for two main reasons. First, the search space given by the ACT-R parameters is supposed to be quite complex, with many local optima and a (μ, λ) selection strategy is better able to avoid local optima. Second, a $(\mu + \lambda)$ selection strategy could be harmful for the autoadaptation mechanism and therefore, a (μ, λ) strategy is preferred [43]

The values of the parameters μ and λ were fixed to $\lambda = 100$ and $\mu = 15$, as an approximate value of $\mu = \lambda/7$ is usually recommended [41,43]. With this configuration, the final optimal values for each parameter were computed as the average value for 10 executions of the evolutionary strategy, which converged for all cases in 32.88 hours on average (note that in the diagnosis of most cognitive impairments, time taken is not the most important constraint). We obtained a global correlation of 0.83 between the children's and the model's vectors.

4.4. Application of machine-learning techniques

As stated in Section 2, the machine-learning approach has been used for the diagnosis of SLI with promising results [26,27]. However, the usefulness of machine-learning techniques depends on the information they are fed. Therefore, as stated above, this stage attempts to demonstrate the importance of both the information source (behavioral variables/cognitive variables) and the techniques to process that information (classical approaches such as correlation or linear discriminant analysis/machine-learning techniques) in the diagnosis and characterization. We performed two different experiments to test the performance of our proposal for both the diagnosis and characterization of SLI. All the experiments were run under RapidMiner [44].

To test the diagnostic performance of our methodology we decided to use four different techniques to compare their performances: a support vector machine (SVM), a naïve Bayes classifier (NB), a decision tree (DT) and a neural network (NN). We chose the reference method used by Simon-Cereijido and Gutiérrez-Clellen [32] and compared it with these four machine-learning techniques. In that work, the authors used a linear discriminant analysis (LDA) with only behavioral variables. We applied the previous steps of the methodology to obtain the best set of parameter values and thus the best model for each of the 48 individuals, and we built four different classifiers for the diagnosis. To evaluate the method's performance, we performed a leave-one-subject-out cross validation (LOOCV) and computed the sensitivity, specificity and the area under curve (AUC). Note that the results reported here for the LDA technique do not exactly match the ones reported by Simon-Cereijido and Gutiérrez-Clellen [32] because we used different evaluation techniques. Those authors validated the model with the same training set and with a test set with only five individuals per class. We decided not to use that evaluation method because the number of individuals per class is too low, and, therefore, a LOOCV method seemed to be more appropriate. To obtain the best parameter configuration for each method, we used the Evolutionary Parameter Optimization method implemented in RapidMiner. For the SVM, we optimized the *kernel gamma* and *C* parameters. For the DT, we optimized the *minimal size for split*, the *maximal depth* and the *confidence* parameters. For the NN, we optimized the *learning rate* and the *momentum* parameters. The NB classifier does not have parameters to optimize.

In terms of characterization, we wanted to check whether the difficulties posed by the heterogeneity of SLI could be solved by using the information related to the underlying mechanisms obtained through our cognitive models. In particular, this experiment has three subobjectives: to confirm the informative value of the cognitive variables, to identify different subgroups inside the SLI profile (if they exist, as some authors point out [4,5]) and to seek clues to support any of the existing theories on SLI. To achieve these objectives, we first applied the hierarchical clustering implementation given by RapidMiner [44], using the observable feature set and the cognitive feature set separately as input. Then, we applied a k-means clustering algorithm to study the centroid values of each of the clusters obtained through the hierarchical cluster. To analyze the importance of the different information sources we built three different feature sets:

Observable feature set. The first feature set consists of only the observable variables captured in the database. The complete list of features is described in Section 4.1. Further details can be found in Simon-Cereijido and Gutiérrez-Clellen [32]

Cognitive feature set. The second feature set consists of only the internal variables of the model (i.e., the parameter values obtained from the individualized computational models). A description of these parameters is given in the previous section.

Complete feature set. The third feature set is a combination of the two previous sets and thus includes both behavioral and cognitive variables.

5. Results and discussion

5.1. Diagnosis

The results obtained with the three feature sets are shown in Table 2. We performed two statistical tests on the performance results: a one-way ANOVA to check whether the differences for each classifier and the reference method are statistically significant, and a one-way ANOVA to check whether the difference for classification performance of each classifier with the different feature sets is statistically significant.

First, we examined the importance of machine learning techniques for the diagnosis support. All the classifiers except the DT performed better ($p < 0.01$) than the baseline method in terms of sensitivity and AUC. It is worth remarking that, in this case, it is more useful to obtain good results in terms of sensitivity because diagnosing an child with impairment as normal is much more problematic than the other way round. Therefore, we can conclude, similar to some previous works [26,27], that the use of machine-learning techniques can be of great interest for the diagnosis of cognitive impairments and, in particular, for the diagnosis of SLI.

The other main objective of this experiment is to show that the diagnosis of cognitive impairments could be improved by using the information obtained from computational cognitive modeling. The results clearly show the importance of that information. We found significant differences ($p < 0.01$) among the three feature sets in terms of sensitivity for each classifier. All the classifiers improved their sensitivity when the model's cognitive variables were used. The most relevant result is that using the complete feature set produced a significant improvement ($p < 0.01$) as compared with the other two feature sets in all cases, in terms of sensitivity, specificity and AUC.

Analyzing the results in greater depth, we found that two normal children were diagnosed by our methodology as having SLI (false positives) by all the classifiers but one (the neural network with the complete feature set). Common to both these cases is the fact that they have a greatly reduced value for the α parameter. As we will see below, a low value in that parameter is found in most SLI

Table 2
Sensitivity (Sensit.), specificity (Specif.) and area under curve (AUC) for the different methods and feature sets. SVM and LDA stand for support vector machine and linear discriminant analysis, respectively.

	Observable			Cognitive			Complete		
	Sensit.	Specif.	AUC	Sensit.	Specif.	AUC	Sensit.	Specif.	AUC
SVM	0.71	0.84	0.82	0.78	0.88	0.85	0.98	0.94	0.95
Naïve Bayes	0.73	0.85	0.80	0.84	0.79	0.82	1	0.92	0.95
Decision tree	0.69	0.75	0.73	0.76	0.80	0.77	0.92	0.89	0.91
Neural net	0.73	0.81	0.78	0.82	0.81	0.81	0.97	1	0.97
LDA	0.63	0.79	0.72	0.77	0.78	0.77	0.89	1	0.90

children. These two children most likely present a low α parameter due to a general learning problem or a simple delay that has nothing to do with SLI. On the other hand, one child with impairment was also diagnosed as normal (false negative) by all the classifiers but one (the naïve Bayes classifier with the complete feature set). The confusing factor in this case is that the deficit seems to be spread over all areas and parameters, as there is no specific parameter that is greatly different from normal children. Both types of error demonstrate once again the difficulty of the task and the existing overlap between SLI children and normal children, not only at the behavioral level.

We conclude that the combination of machine-learning techniques with the information obtained through computational cognitive modeling could be a helpful methodology to support the diagnosis of SLI. All the classifiers tested except the DT reached sensitivity, specificity and AUC values above 90%, reaching 100% in some cases. Moreover, one interesting feature of our proposal is that, given an individual, our methodology does not only suggest a binary diagnosis in terms of "impaired" or "non-impaired". Our diagnosis methodology gives clues (the parameter values) as to which underlying mechanisms are impaired and to what extent they are impaired in each individual. This information is very useful for the design of individualized therapies that would most likely improve the recovery process.

5.2. Characterization

The dendrograms shown in Figs. 1 and 2 show the results of hierarchical clustering with the observable feature set and with the cognitive feature set respectively. We could not find any well-defined group using the observable feature set. However, with the cognitive feature set, we obtained three very well differentiated main clusters (the minimum distance between members of different groups is 3.66, while the maximum distance between members of the same group is 1.76). These results confirm the first two objectives of this experiment. First, while the clustering obtained by using the observable variables provides no information, the clustering obtained by using the cognitive feature set shows three very well differentiated groups. Therefore, we can conclude that the internal variables are important not only for classification but also for the characterization process. The second objective consisted of finding different subgroups (if they exist) inside the SLI profile. Obtaining three clearly separate groups suggests that it is difficult to find a unique general characterization for SLI. However, the great cohesion of those clusters shows that, despite the highly heterogeneous profile shown by SLI, some subgroups can be identified. In other words, heterogeneity at the behavioral level is extended, up to a point, to the underlying mechanisms. However at the cognitive level, we can find patterns and subgroups that are very useful for characterization of the impairment.

The third objective consisted of trying to find evidence to support any of the theories that try to account for the behavioral profile of SLI. To determine to what extent the three subgroups found match the existing theories, we applied a k-means clustering algorithm. Again, we used the implementation given by RapidMiner and the parameter values for the 24 SLI individuals in the database obtained during the subject modeling stage of the methodology.

The coordinates of the centroids of the three clusters are shown in Table 3. Each cluster presents statistically significant differences from the healthy group for different parameters. Some statistically significant differences were found for the observable variables. However, their predictive power is much lower, as we showed in the previous section. Fig. 3 shows the normalized values of those coordinates and the mean parameter values for the 24 TD children in the database. Thus, we can compare which areas are impaired in each cluster.

As can be observed, the three clusters obtained can be related to the three theories presented in Section 2.

Cluster 1 - PDH. This cluster shows differences in procedural memory parameters. Both the α and EGS parameters show significant differences. The α parameter controls how fast new rules are acquired, and the EGS parameter controls the noise in procedural memory. Differences in these two parameters suggest a general deficit in procedural memory, not related solely to language.

Fig. 1. Hierarchical clustering of individuals with specific language impairment with the observable feature set.

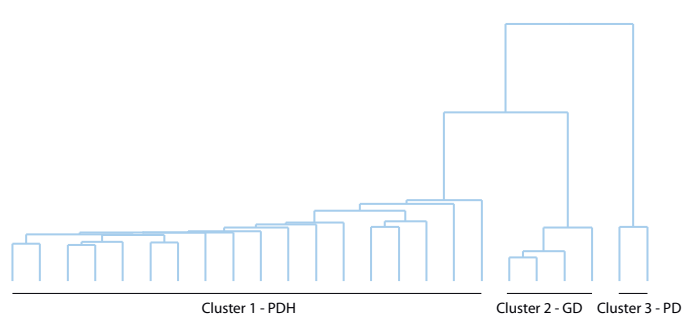

Fig. 2. Hierarchical clustering of individuals with specific language impairment with the cognitive feature set.

Table 3
Centroid values: mean values and standard deviations. The abbreviations used in the table are the following: EGS: *expected gain S*, RT: *retrieval threshold*, ANS: *activation noise S*, BLL: *base level learning*, C-PM: *conjugation partial matching*, MTA-PM: *mood-time-aspect partial matching*, NP-PM: *number-person partial matching* and Ph-PM: *phonological partial matching*.

		Cluster 1	Cluster 2	Cluster 3	Healthy
rocedural memory	α	3.7e−4** (1.6e−4)	1.0e−3 (2.5e−4)	8.4e−4 (1.3e−4)*	1.1e−3 (2.4e−4)
	EGS	0.23** (0.043)	0.281** (0.016)	0.190 (0.003)	0.1946 (0.016)
Declarative memory	RT	−0.114 (0.044)	−0.112 (0.045)	−0.092 (0.007)	−0.097 (0.022)
	ANS	0.283 (0.036)	0.311** (0.007)	0.255 (0.201)	0.276 (0.014)
	BLL	0.314 (0.014)	0.277* (0.023)	0.315 (0.014)	0.315 (0.016)
Grammatical processing	C-PM	−6.51 (0.39)	−2.97** (0.10)	−6.17** (0.02)	−6.51 (0.39)
	MTA-PM	−2.23 (0.39)	−0.53** (0.08)	−2.16 (0.74)	−2.31 (0.39)
	NP-PM	−2.22* (0.74)	−1.08** (0.10)	−2.52 (0.12)	−2.63 (0.12)
	γ_m	0.133 (0.082)	0.204 (0.149)	0.228** (0.005)	0.1128 (0.031)
	γ_m-decay	16.36 (1.61)	16.11 (0.85)	8.065** (0.38)	17.09 (1.42)
Phonological processing	Ph-PM	−3.25* (0.62)	−3.94* (0.51)	−4.05 (0.48)	−5.62 (0.34)
	γ_{ph}	0.081 (0.064)	0.106 (0.012)	0.163** (0.037)	0.093 (0.05)
	γ_{ph}-decay	12.463 (1.22)	14.563 (0.93)	13.654** (0.67)	16.854 (0.32)

* Significant difference with the corresponding value on healthy children using a two tailed *t*-test ($p < 0.05$).
** Significant difference with the corresponding value on healthy children using a two tailed *t*-test ($p < 0.01$).

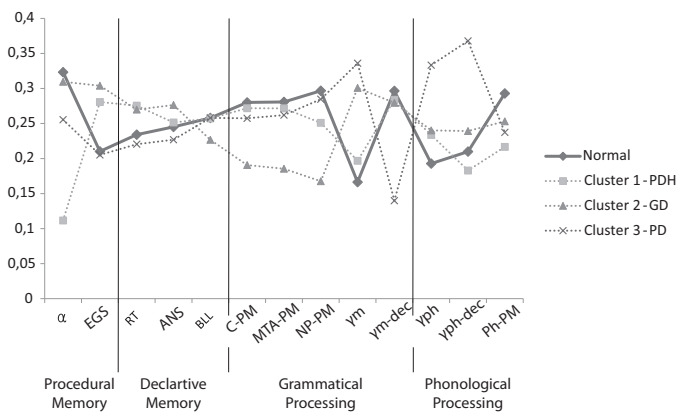

Fig. 3. Normalized values for the centroids of the three clusters obtained for the individuals with specific language impairment.

Cluster 2 - GD. This group clearly shows great problems in the processing of grammatical features such as mood, time, aspect, number, person or conjugation. Partial matching parameters related to grammatical features have values much lower than the normal ones, which shows that these children have difficulties in differentiating and processing those features. However, the perception of these grammatical features is not especially impaired. Morphological noise (γ_m) is higher than the normal case, but the decay of that noise (γ_m-decay) is similar to the value obtained for normal children. The characteristics of this cluster resemble the ones proposed by the GD theories, which argue that SLI is produced by problems in the processing of grammatical features.

Cluster 3 - PD. The main differences from the normal group are detected in the morphological and phonological noise parameters (γ_m and γ_{ph} respectively). This fact suggests a general deficit in information processing. Therefore, this group roughly matches the PD theories. However, those theories group many different approaches, and a finer analysis should be performed to examine approaches that suggest deficits in different processing elements, such as phonological processing or working memory processing.

The most numerous group by far is the first one (cluster 1). A procedural deficit is shared by 18 out of 24 individuals. Moreover, the other two groups share some problems in procedural memory. The PD group has a reduced α parameter compared to normal children, and the GD group presents high procedural noise (EGS parameter). The results show that the three main theories could explain some individual cases, but our methodology suggests that all cases share some type of procedural deficit. Accordingly, PDH seems to be the most plausible explanation for SLI. Nonetheless, the heterogeneous profile observed at the behavioral level is also observed at the cognitive level. However, it seems easier to find patterns in the underlying mechanisms at the cognitive level and, as a consequence, to define subgroups. Given this heterogeneity at all levels, it might be important to focus not on searching for a general explanation of SLI but on the individual differences that each patient presents. This fact shows again the importance of the design of individualized methods for the diagnosis and characterization of impairment, such as the method presented in this work.

6. Conclusions and future work

In this paper, we propose a new approach for the characterization and diagnosis of cognitive impairments. Given the limitations of the methods that rely on behavioral variables, our methodology consists of using information related to the cognitive mechanisms underlying normal and impaired behavior. One of the key points of our methodology is the use of computational cognitive modeling to obtain information on these underlying mechanisms. In addition, we prove the predictive power of this information by using supervised machine-learning techniques, which have been tested with promising results for similar tasks. We evaluated the proposed methodology with a dataset of normal and SLI Spanish-speaking children for a narrative production task. However, the methodology presented could easily be extended to other languages and even to other cognitive impairments, not necessarily related to language.

In terms of diagnosis, the results showed that the combination of machine learning techniques with the information obtained through computational cognitive modeling could be a helpful methodology to support the diagnosis of SLI. All the classifiers tested except the DT achieved sensitivity, specificity and AUC values above 90%, reaching 100% in some cases. It is worth noting that these results were achieved with a database of only 48 children and, therefore, this approach should be tested with larger samples to confirm these results. Another important feature of the proposed methodology is the high information content of the diagnosis it suggests. In a given individual, our methodology does not simply suggest a binary diagnosis in terms of “impaired” or “non-impaired”. This diagnosis methodology gives clues as to which underlying mechanisms are impaired and to what extent they are impaired in each individual. This information is very useful for the

design of individualized therapies that would most likely improve the recovery process.

In terms of characterization, the results seem to confirm the PDH. All individuals showed an impaired procedural memory. Moreover, the results show the importance of the proposed methodology for the characterization of SLI. At the behavioral level, it is difficult to correctly differentiate among different subgroups. However, at the cognitive level, it is easier to find clear differences despite the heterogeneity.

Therefore, the three salient points of our approach are as follows:

- The informative capacity and predictive power of the features obtained from computational cognitive models are higher than the observable variable features
- Machine-learning techniques improve the classification results obtained with the classical approaches, such as linear discrimination analysis.
- Clustering techniques combined with the information obtained from the model improve the characterization of the impairment, detect subgroups and support one of the existing theories (PDH).

Our future work will focus on two primary issues. First, we aim to test the generality of the methodology by using it to diagnose SLI in other languages and other impairments not related to language. Second, we aim to close the loop with psycholinguists and therapists to confirm the results shown in this paper. We also plan to extend the work presented in this paper by using larger databases and by studying and optimizing alternative machine-learning techniques.

Acknowledgements

We are grateful to Gabriela Simon-Cereijido and Vera F. Gutiérrez-Clellen for making available the full contents of their study with Spanish-speaking children. We also thank the JAE-Pre fellowship program of the Spanish National Research Council (CSIC) and the financial support of FGCSIC, Obra Social la Caixa, CSIC and the project PIE 201050E087.

References

[1] McLean L, Woods-Cripe J. The effectiveness of early intervention for children with communicative disorders. Baltimore, MD: Paul Brooks Publishing; 1997. p. 349–428.
[2] Leonard L. Children with specific language impairment. Cambridge, MA: MIT Press; 1998.
[3] Conti-Ramsden G, Botting N. Classification of children with specific language impairment: longitudinal considerations. J Speech Lang Hear Res 1999;42(5):1195–204.
[4] Bishop D. Pragmatic language impairment: a correlate of SLI, a distinct subgroup or part of the autistic continuum? Speech and language impairments in children: causes, characteristics, intervention and outcome; 2000. p. 99–113.
[5] Gopnik M, Crago M. Familial aggregation of a developmental language disorder. Cognition 1991;39(1):1–50.
[6] Tirosh E, Cohen A. Language deficit with attention-deficit disorder: a prevalent comorbidity. J Child Neurol 1998;13(10):493–7.
[7] Hill E. Non-specific nature of specific language impairment: a review of the literature with regard to concomintant motor impairments. Int J Lang Commun Disord 2001;36(2):149–71.
[8] Montgomery J. Working memory and comprehension in children with specific language impairment: what we know so far. J Commun Disord 2003;36:221–31.
[9] Rice M, Wexler K, Cleave P. Specific language impairment as a period of extended optional infinitive. J Speech Hear Res 1995;38(4):850–63.
[10] Wexler K. Optional infinitives. New York: Cambridge University Press; 1994. p. 305–51 [Chap. 14].
[11] Van der Lely H. Canonical linking rules: forward versus reverse linking in normally developing and specifically language-impaired children. Cognition 1994;51(1):29–72.
[12] Bishop D, Adams C, Rosen S. Resistance of grammatical impairment to computerized comprehension training in children with specific and non-specific language impairments. Int J Lang Commun Disord 2006;41(1):19–40.
[13] Leonard L, McGregor K, Allen G. Grammatical morphology and speech perception in children with specific language impairment. J Speech Lang Hear Res 1992;35(5):1076–85.
[14] Leonard L, Eyer JA, Bedore L, Grela B. Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. J Speech Lang Hear Res 1997;40(4):741–53.
[15] Van der Lely H, Ullman M. Past tense morphology in specifically language impaired and normally developing children. Lang Cogn Process 2001;16(2–3):177–217.
[16] Gathercole S, Baddeley A. Phonological memory deficits in language disordered children: is there a causal connection? J Mem Lang 1990;29(3):336–60.
[17] Ullman M. A neurocognitive perspective on language: the declarative/procedural model. Nat Rev Neurosci 2001;2(10):717–26.
[18] Ullman M, Pierpont E. Specific language impairment is not specific to language: the procedural deficit hypothesis. Cortex 2005;41(3):399–433.
[19] Thomas M, Karmiloff-Smith A. Can developmental disorders reveal the component parts of the human language faculty? Lang Learn Dev 2005;1(1):65–92.
[20] Hsu H, Bishop D. Grammatical difficulties in children with specific language impairment: is learning deficient? Hum Dev 2010;53(5):264–77.
[21] Spaulding T, Plante E, Farinella K. Eligibility criteria for language impairment: is the low end of normal always appropriate? Lang Speech Hear Serv School 2006;37(1):61–72.
[22] Prud'hommeaux E, Roark B. Alignment of spoken narratives for automated neuropsychological assessment. In: Proceedings of the IEEE workshop on automatic speech recognition and understanding (ASRU). Hawaii: IEEE; 2011. p. 484–9.
[23] Prud'hommeaux E, Rouhizadeh M. Automatic detection of pragmatic deficits in children with autism. In: Proceedings of the 3rd workshop on child, computer and interaction (WOCCI 2012). 2012.
[24] Georgopoulos V, Malandraki G, Stylios C. A fuzzy cognitive map approach to differential diagnosis of specific language impairment. Artif Intell Med 2003;29(3):261–78.
[25] Solorio T, Liu Y. Using language models to identify language impairment in Spanish-English bilingual children. In: Demner-Fushman, editor. Proceedings of the workshop on current trends in biomedical natural language processing. Stroudsburg, PA: Association for Computational Linguistics; 2008. p. 116–7.
[26] Gabani K, Sherman M, Solorio T, Liu Y, Bedore L, Peña E. A corpus-based approach for the prediction of language impairment in monolingual English and Spanish-English bilingual children. In: Proceedings of human language technologies: the 2009 annual conference of the North American chapter of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2009. p. 46–55.
[27] Gabani K, Solorio T, Liu Y, Hassanali K, Dollaghan C. Exploring a corpus-based approach for detecting language impairment in monolingual English-speaking children. Artif Intell Med 2011;53(3):161–70.
[28] McCaulley R, Fey M. Treatment of language disorders in children. Baltimore, MD: Paul Brookes Publishing; 2006.
[29] Bedore L, Leonard L. Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. Appl Psycholinguist 2005;26(2):195–225.
[30] Restrepo M, Gutiérrez-Clellen V. Grammatical impairments in Spanish–English bilingual children. Baltimore, MD: Brookes; 2004. p. 213–34.
[31] Sanz-Torrent M, Serrat E, Andreu L, Serra M. Verb morphology in Catalan and Spanish in children with specific language impairment: a developmental study. Clin Linguist Phon 2008;22(6):459–74.
[32] Simon-Cereijido G, Gutiérrez-Clellen V. Spontaneous language markers of Spanish language impairment. Appl Psycholinguist 2007;28(2):317–39.
[33] Mayer M. Frog goes to dinner. New York: Dial Press; 1974.
[34] Mayer M. Frog on his own. New York: Dial Press; 1973.
[35] Miller JF, Chapman R. Systematic analysis of language transcripts, Tech. rep. Madison, WI: Language Analysis Laboratory, University of Wisconsin; 2000.
[36] Anderson J. How can the human mind occur in the physical universe? New York: Oxford University Press; 2007.
[37] Oliva J, Serrano JI, Del Castillo MD, Iglesias A. Cognitive modeling of the acquisition of a highly inflected verbal system. In: Salvucci D, Gunzelmann G, editors. Proceedings of the 10th international conference on cognitive modeling. Philadelphia, PA: Drexel University; 2010. p. 181–6.
[38] Thomas M, Karmiloff-Smith A. Modeling language acquisition in atypical phenotypes. Psychol Rev 2003;110(4):647–82.
[39] Rivera S, Bates E, Orozco-Figueroa A, Wicha N. Spoken verb processing in Spanish: an analysis using a new online resource. Appl Psycholinguist 2010;31(1):29–57.
[40] Taatgen N, Anderson J. Why do children learn to say "broke"? A model of learning the past tense without feedback. Cognition 2002;86(2):123–55.
[41] Beyer H, Schwefel H. Evolution strategies: a comprehensive introduction. Nat Comput 2002;1(1):3–52.
[42] Wong T, Cokely E, Schooler L. An online database of ACT-R parameters: towards a transparent community-based approach to model development. In: Proceedings of ICCM – 2010 tenth international conference on cognitive modeling. Philadelphia, PA: Drexel University; 2010. p. 282–6.
[43] Eiben A, Smith J. Introduction to evolutionary computation. Berlin: Springer; 2002.
[44] Mierswa I, Wurst M, Klinkenberg R, Scholz M, Euler T. Yale Rapid prototyping for complex data mining tasks. In: KDD'06: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY: ACM; 2006. p. 935–40.