Inferring Learning Strategies from Students’ Errors

Lila Ghemri
Department of Computer Science
Texas Southern University

Abstract - Student models in an Intelligent Tutoring System (ITS) aim at representing both correct and incorrect knowledge of the student in a particular domain of learning. This data serves as the basis upon which the Tutor can dispense remedial feedback. This paper presents a framework in which learners’ errors are diagnosed, classified and linked to cognitive processes associated with learning. The cognitive processes we identify are those of generalization, misattribution and ignorance. Our system applies to the domain of Second Language Acquisition and our approach uses Chomsky’s linguistic theory of Principles and Parameters. We particularly concentrate on the acquisition of the Phrase Structure component.

Keywords: ITS, L2 learning, Diagnostic, Feedback.

1. Introduction
Abundant literature on learners misconceptions, describe errors that learners make in their specific domain of learning and the relevant feedback to be provided. However in language learning, the usefulness of feedback or correction has long been controversial and is still the object of a heated debate between linguists and teachers. In Second Language (L2) learning in particular, each of the numerous teaching strategies and pedagogical methods that have been developed, advocates a different way to respond to learners’ misconceptions, from preventing them from happening [8], to ignoring them [4] or even encouraging them [7]. Within the tradition of generative grammars, first language acquisition has been equated with the provision of positive evidence, or correct instances of the language. It is claimed that this should be sufficient for language learners who would have then to work out by themselves the rules of the language. This result has been carried over for L2 acquisition as shown in [12], [13], [14], [15], [16] and [17].

In Intelligent Computer Aided Language Learning (ICALL) systems, in which a learner can only interact with a computer program, feedback seems indispensable; not only to correct the learner but also to minimize his or her frustration and keep them interested and motivated in their learning endeavor. Furthermore, current research in the field has shown that cognitive approaches to diagnosis yield a more accurate representation of the learner’s knowledge and a more valid response to their errors ([6], [9], and [11]).

3. The Principles and Parameters (P&P) Theory of Language

The Principles and Parameters theory, first introduced in [2] and developed further in [3]; defines subsystems of principles which govern the structures of the grammar of any language. The principles serve to reduce the number of rules in the grammar and replace them by a principled account on how these rules can be generated. The problem of how this system of principles leads to such a diversity of languages introduces the concept of Parameters. Parameters account for the differences among languages; they explain why a given construction is correct in some languages, whereas it is considered incorrect in some others.
3.1 Parameters and L2 Learning:

In Second Language acquisition, the learner already possesses the grammar rules of their native language or L1. Acquiring the L2 grammar is thus seen as adding new rules to the set of rules already known by the learner. Within the P&P theory, this acquisition of new rules is explained by the setting of parameter values pertaining to the Second Language.

3.2 Acquisition of Grammar Rules:

Central to the acquisition of a language is its grammar or the phrase structure component. The phrase structure represents an internalized set of grammar rules for a given language. In the Principle and Parameter framework, the phrase structure theory or X-bar theory brings out what is common in the structure of phrases. According to X-bar theory, all phrases are headed by a lexical head. The lexical head is a zero projection (X0). Two further levels of projection are stipulated: 1) complements and potential adjuncts combine with X0 to form X1; 2) specifiers combine with the topmost X1 to form a maximal projection or X2. The following underspecified set of rules is obtained as a result:

\[ X^2 \rightarrow \text{Specifier [Adjuncts]} \ X^1 \]
\[ X^1 \rightarrow \text{Complement} \ X^0 \]

3.3 Directionality Parameters:

The specialization of the X-bar rules to a given language is function of the values of some parameters that mostly relate to directionality. Directionality manifests itself in the position that some phrase structure elements take with regards to one another. The parametric variation in value yields variations in the order of phrase structure components, which result in languages being SVO (subject-verb-object) such as English, or VSO, (verb-subject-object) such as Arabic, etc...

In the present work, following Briscoe [1], we assume that directionality involves the following parameters: spec_head, comp_head, genitive, modifier_head, subjdir, objdir.

Further, we assume that each of these parameters can take a value 0 for right with regard to the head, 1 for left with regard to the head, or can be unset (-1). When a parameter is unset, it is assumed that it has the Universal Grammar (UG) setting. For English, values associated with each of these parameters are (1,0,1,1,1,0). In the case of French, the set of parameter values are (1,0,0,1,1,0) and so on.

A grammar G of a given language L takes the neutral categories of X-bar (UG setting), then applies the values of each parameter to generate language specific rules. Consequently a set of parameter values defines a specific grammar. These language specific rules are obtained through the operations of substitution and adjunction and also through the specialization of the lexical category or head.

3.3 Specializing the Phrase-structure component:

Applying the principles stated above, we define the parser of a given language L as

\[ \text{Parser}(L, \text{Rules}) = (X\text{-bar}(UG), \text{Settings} (\text{Parameters}(L))) \]

For example, in the case of English, this procedure will generate the following (restricted) Phrase-structure rules:

\[ S^2 \rightarrow N^2 S^1 \]
\[ S^1 \rightarrow \text{Infl} \ V^2 \]
\[ N^2 \rightarrow \text{Spec} \ N^1 \]
\[ N^1 \rightarrow \text{Adj} \ N^0 \]
\[ N^1 \rightarrow N^2, s \ N^1 \]
\[ V^2 \rightarrow V^1 N^2 \]
\[ V^1 \rightarrow \text{Adv} \ V^0 \]

This parsing system is fully described in [5].

4. Diagnosis and Learning Strategy

We present a framework for diagnosing learners’ misconceptions in L2 acquisition within the Principles and Parameters theory. Particularly, we will consider grammatical errors that stem from parametric variation. These errors occur when the learner uses a parameter value different from the one that the L2 allows. Our system infers the
learner’s settings of these parameters based on the sentences he/she produces.

In the diagnosis process, we used the approach described in [10] and adapted it to the domain of L2 learning. This approach allowed us to not only diagnose errors; but also the cognitive processes that caused these errors. The processes are mostly domain independent and can be: generalization; misattribution; or ignorance.

Misconceptions that originate from an incorrect parameter setting will manifest themselves in various ways:

1. Value of the parameter is still unset, or has the UG setting.
2. Value of a parameter is generalized across languages (transfer from L1).
3. Acquisition of the rules related to a parameter value is incomplete.
4. Rules related to a given parameter are generalized across lexical categories.

Consequently, if a misconception relates to a given parameter; the form that the feedback takes depends not only on the parameter but also on the nature of the misconception.

For example, case 1 and case 3 require opposing feedback although they could both relate to the same parameter.

5. Computational System

In what follows, we present a description of the algorithms used to identify misconceptions and label them in terms of learning strategies. We shall also present the algorithm that computes the required feedback.

In order to diagnose errors, it is necessary that the system possesses a model of correctness so that comparisons could be made with regard to it. This model of correctness is the set of the values of parameters of the Second Language being acquired.

5.1 Notation:

Let A be a learner and L2 be the target language A is attempting to acquire. The system computes a representation of A’s knowledge of L2 and stores it in a database D. This representation consists of parameters, their values and their corresponding phrase structure rules.

We define a learning function $lf$ that realizes a mapping between a parameter $p$ and its corresponding phrase structure rules in $D$.

The system also possesses a representation of the language L2 in terms of parameters, their settings and the corresponding phrase structure rules, this information is contained in a database $Z$.

A function $e$ realizes the mapping between a parameter $p$ in L2 and its corresponding rules. Ideally, $Z$ and $D$ should be identical. When they differ, it is an indication that the learner has a misconception.

$$\text{A} : \text{a learner}$$

$$\text{L2} : \text{a second language to be acquired}$$

$$\text{D} : \text{a database of parameters and phrase structure rules representing A’s knowledge of L2.}$$

$$\text{p} : \text{a parameter}$$

$$\text{P} : \text{the set of all parameters p}$$

$$\text{t}_p : \text{phrase-structure rules resulting from a value of the parameter p.}$$

$$\text{T} : \text{The set of all phrase-structure rules.}$$

$lf$: a mapping defined by $A$ such that:

$$lf_A(p) = q = (t_1, t_2, \ldots, t_m)$$

$Q_A : \text{the set of all q tuples}$

$B_A(t, p) : \text{true if A uses the rule t for the parameter p.}$

$e_{L2}$: a mapping from $P$ to a $T$ such that:

for all $p$ belonging to $P$,

$$e_{L2}(p) = z = (s_1, s_2, \ldots, s_n), \text{with each s being a phrase-structure rule resulting from the L2 setting of parameter p.}$$

$Z_{L2} = \text{the set of all z tuples.}$

5.2 Computing L2 Representations:

Diagnosing misconceptions is possible once we have the set $Q_A$, which is the set of Phrase Structure rules that a learner has as their representation of the language L2 and the set $Z_{L2}$,
which comprises the Phrase Structure rules associated with L2. ZL2 represents the correct skills against which the learner’s representation of L2 will be compared.

a. Step 1: Computing QA(p):
   \[ q = \emptyset \]
   For all \( t \) in \( D_L \)
   if \( B_A(t, p) \) then collect \( t \) in \( q \)
   endif
   endfor
   QA(p) = q
   end

b. Step 2: Computing ZL2,
   begin
   \( x = \emptyset \)
   For all \( t \) in \( L_2 \)
   if \( e_{L_2}^{-1}(t) = p \) then collect \( t \) in \( x \)
   endif
   endfor
   ZL2(p) = x

5.3 Classifying misconceptions and Identifying Learning Strategies:

This algorithm provides a classification of the misconceptions. The comparison between QA(p) and ZL2(p) will fall into six cases:

begin:
1. if \( QA(p) = ZL2(p) \) then there is no misconception involving \( p \) in \( D \).
2. if \( QA(p) = \emptyset \) and \( ZL2(p) \neq \emptyset \) then there is no representation of \( p \) in \( D \).
3. if \( QA(p) \neq \emptyset \) and \( ZL2(p) = \emptyset \) then there is an instance of generalization involving \( p \) in \( D \).
4. if \( QA(p) \cap ZL2(p) = \emptyset \) then there is no representation of the rules of \( p \) in \( D \).
5. if \( QA(p) \subset ZL2(p) \) then there is at least a rule of \( p \) with no representation in \( D \).
6. if \( ZL2(p) \subset QA(p) \) then there is an instance of generalization of a rule of \( p \) in \( D \).

5.4 Computing the feedback:

We shall now examine how each case of misconception is treated and how feedback, denoted \( F(p) \), is provided. The content of the feedback is dependent on the type of the misconception. Some misconceptions require the provision of positive evidence while others require negative evidence -correction- or a combination of both.

We define two functions \( E_{pos} \) and \( E_{neg} \) to denote the type of feedback; \( E_{pos} \) for feedback providing positive evidence, and \( E_{neg} \) for feedback providing correction.

When the nature of the misconception is related to an incomplete knowledge, the feedback should provide positive evidence of the missing items. When the misconception relates to a generalization of a piece of knowledge, then this generalization should be refuted through correction or negative evidence.

The case of misattribution requires both types of evidence, in that the incorrect knowledge must be refuted together with presenting the correct knowledge.

In the first three cases, the feedback is computed as: \( F(p) = ZL2(p) \setminus QA(p) \)

Case 1: This is the ideal case, where the parameter and its related tasks have been correctly acquired. In this case, the formula correctly computes \( F(p) = \emptyset \).

Case 2 and 5: These two cases involve missing knowledge in the learner’s representation of \( p \) in \( L_2 \). They require that evidence of the existence of the missing items be provided. This is a case requiring positive evidence. The way the feedback is provided is: \( E_{pos}(F(p)) \)

Case 2:
\[ QA(p) = \emptyset; \text{ the learner has to acquire all of } ZL2(p) \text{ and the formula correctly computes } F(p) = ZL2(p) \]

Case 5:
The learner has partially acquired some rules related to the parameter \( p \) and needs to acquire the missing items from \( ZL2(p) \). The formula correctly computes this set.

Case 3 and 6:
Both of these cases involve generalizations in the learner’s representation of \( p \) in \( L_2 \). Case 3 relates
to the incorrect acquisition of parameter p, whereas case 6 relates to the incorrect acquisition of a rule of p.
Correcting generalizations requires the provision of negative evidence, because the generalization must be refuted. For both cases, the items to be refuted are:
\[ F(p) = Q_A(p) \setminus Z_{L2}(p) \]
The feedback has to be provided negatively:
\[ E_{neg}(F(p)) \]

**Case 3:**
\[ Z_{L2}(p) = \emptyset \] and all tasks present in \( Q_A(p) \) must be refuted as computed by the function.

**Case 6:**
Only the tasks that are contained in \( Q_A(p) \) and not \( Z_{L2}(p) \) should be refuted.

**Case 4:**
This case is an instance of misattribution of rules to a parameter, in that the learner maps a parameter with the wrong rules. This case requires positive evidence to show the correct rules and negative evidence to refute the incorrect ones. The feedback is:
\[ F(p) = Q_A(p) \cup Z_{L2}(p) \]
The way to provide feedback is:
\[ E_{pos}(Z_{L2}(p)) \land E_{neg}(Q_A(p)) \]
That is providing positive evidence for the missing rules in \( Z_{L2}(p) \) and negative evidence for the generalized rules in \( Q_A(p) \).

Based on these rules, the program computes the set \( Q_A \) as: \( Q_A(subjdir) = \{0\} \), \( Q_A(objdir) = \{1\} \)
Now it will compute the corresponding \( Z_{L2}(subjdir) \) and \( Z_{L2}(objdir) \) and we obtain
\[ Z_{L2}(subjdir) = \{1\} \text{ and } Z_{L2}(objdir) = \{1\} \]

Turning now to diagnosis and feedback, and taking one parameter at a time, we apply our rules and obtain:

1. \( Q_A(objdir) = Z_{L2}(objdir) \), there is no misconception and the feedback is correctly computed as:
\[ F(objdir) = Q_A(objdir) \setminus Z_{L2}(objdir) = \emptyset \]

2. \( Q_A(subjdir) = \{0\} \),
\( Z_{L2}(subjdir) = \{1\} \),
\( Q_A(subjdir) \cap Z_{L2}(subjdir) = \{0\} \cap \{1\} = \emptyset \).
This case falls into case 4 of our algorithm, which is a case of misattribution of rules of the parameter subjdir. The feedback is:
\[ F(subjdir) = Q_A(subjdir) \cup Z_{L2}(subjdir) = \{0, 1\}. \]
The way to provide feedback is:
\[ E_{pos}(1) \land E_{neg}(0) \] that is refuting the verb-subject word order and providing positive evidence for subject-verb word order.

5.5 A working example:

We now present an example of how this system works. Let us assume that the user enters the following sentence (cited in [7])

- Eat camel grass

Considering the directionality parameters subjdir, objdir and the corresponding parsing rules:

\[ S^2 \rightarrow N^2 S^1 \text{ (subjdir)} \]
\[ V^2 \rightarrow V^1 N^2 \text{ (objdir)} \]

We have presented a computational framework for diagnosing word order errors using the principles and parameters paradigm. This framework diagnoses errors and classifies them in terms of cognitive processes such as generalization, misattribution, etc. This system also computes the feedback that is required for addressing the learner’s misconceptions.

It is worth noting that this system can be applied to any L2 language once its lexicon, and specific parameter settings have been provided.
Our next step is to conduct extensive testing of the system and to expand its coverage.
References:


