HAL open science

A MDL-based Model of Gender Knowledge Acquisition

Harmony Marchal, Benoît Lemaire, Maryse Bianco, Philippe Dessus

To cite this version:

Harmony Marchal, Benoît Lemaire, Maryse Bianco, Philippe Dessus. A MDL-based Model of Gender Knowledge Acquisition. 12th Conference on Natural Language Learning (CoNLL-2008), Aug 2008, Manchester, United Kingdom. pp.73-80. hal-00311720

HAL Id: hal-00311720

https://hal.science/hal-00311720

Submitted on 20 Aug 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A MDL-based Model of Gender Knowledge Acquisition

Harmony Marchal[1], Benoît Lemaire[2], Maryse Bianco[1], and Philippe Dessus[1]
[1]L.S.E. and [2]Laboratoire TIMC-IMAG
University of Grenoble, FRANCE
<first name>.<last name>@upmf-grenoble.fr

Abstract

This paper presents an iterative model of knowledge acquisition of gender information associated with word endings in French. Gender knowledge is represented as a set of rules containing exceptions. Our model takes noun-gender pairs as input and constantly maintains a list of rules and exceptions which is both coherent with the input data and minimal with respect to a minimum description length criterion. This model was compared to human data at various ages and showed a good fit. We also compared the kind of rules discovered by the model with rules usually extracted by linguists and found interesting discrepancies.

1 Introduction

In several languages, nouns have a gender. In French, nouns are either masculine or feminine. For example, you should say *le camion* (the truck) but *la voiture* (the car). Gender assignment in French can be performed using two kinds of information. Firstly, *lexical information*, related to the co-occurring words (e.g., articles, adjectives) which most of times marks gender unambiguously. Secondly, *sublexical information*, especially noun-endings, are pretty good predictors of their grammatical gender (e.g., almost all nouns endings in *–age* are masculine). Several word endings can be used to reliably predict gender of new words but this kind of rules is never explicitly taught to children: they have to implicitly learn that knowledge from exposure to noun-gender pairs. It turns out that children as young as 3 already constructed some of these rules, which can be observed by testing them on pseudo-words (Karmiloff-Smith, 1979).

This paper presents an iterative model of the way children may acquire this gender knowledge. Its input is a large random sequence of noun-gender pairs following the distribution of word frequency at a given age. It is supposed to represent the words children are exposed to. The model constantly maintains a list of rules and exceptions both coherent with the input data and minimal with respect to an information theory criterion. This model was compared to human data at various ages and showed a good fit. We also compared the kind of rules discovered by the model with rules usually extracted by linguists and found interesting discrepancies.

2 Principle of Simplicity

Gender knowledge is learned from examples. Children are exposed to thousands of nouns which are most of the time accompanied with a gender clue because of their corresponding determiner or adjective. For instance, when hearing "*ta poussette est derrière le fauteuil*" [your stroller is behind the armchair], a child knows that *poussette* is feminine because of the feminine possessive determiner *ta,* and that *fauteuil* is masculine because of the masculine determiner *le*. After processing thousands of such noun/gender pairs, children acquired some gender knowledge which allows them to predict the gender of pseudo-words (Marchal et al., 2007; Meunier et al., 2008). This knowledge is largely dependent on the end of the words since the endings of many nouns in French are associated more often with one gender than the other (Holmes & Segui, 2004). For instance children would predict that pseudo-words such as *limette* or *mossette* are rather feminine words although they never heard them before. It means that they should have constructed a rule-like knowledge saying that "*words ending in* -ette *are rather feminine*". Or maybe it is "*words ending in* -te *are rather feminine*" or even "*words ending in* -e

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (http://creativecommons.org/licenses/by-nc-sa/3.0/). Some rights reserved.

CoNLL 2008: Proceedings of the 12th Conference on Computational Natural Language Learning, pages 73–80
Manchester, August 2008

are rather feminine"… Actually, there are many ways to structure this knowledge, especially because this kind of rule generally has exceptions. Let us take an example. Consider the following words and their gender (masculine or feminine): *barrage* [weir] (m), *image* [image] (f), *courage* [courage] (m), *plage* [beach] (f), *étage* [floor] (m), *garage* [garage] (m), *collage* [collage] (m). Several rules could be constructed from this data:

(1) words ending in *-age* are masculine except *image* and *plage*;

(2) words ending in *-age* are feminine except *barrage*, *courage*, *étage*, *garage* and *collage*;

(3) words ending in *-age* are feminine except words ending in *-rage*, *étage* and *collage*.

The latter is an example of a rule whose exceptions may themselves contain rules. The question is to know which rules may be constructed and used by children, and which cognitive mechanisms may lead to the construction of such rules. In order to investigate that issue, we relied on the assumption that children minds obey a principle of simplicity.

This principle is a cognitive implementation of the Occam's razor, saying that one should choose the simplest hypothesis consistent with the data. This idea has already been used in the field of concept learning where it would dictate that we induce the simplest category consistent with the observed examples—the most parsimonious generalization available (Feldman, 2003). Chater & Vitányi (2003) view it as a unifying principle in cognitive science to solve the problem of induction in which infinitely many patterns are compatible with any finite set of data. They assume "that the learner chooses the underlying theory of the probabilistic structure of the language that provides the simplest explanation of the history of linguistic input to which the learner has been exposed." (Chater & Vitányi, 2007).

One way to implement this idea is to consider that the simplest description of a hypothesis is the shortest one. Without considering frequency of the rule usage, rule 1 in the previous example seems intuitively more likely to be used by humans because it is the shortest.

Intuitively, counting the number of characters of each hypothesis could seem a good method but it is better to choose the most compact representation (Chater, 1999). More important, the choice should also depend on the frequency of rule usage: the description length of a rule that would be frequently used should not be counted like a seldom used rule. For instance, rule 2 could be a more appropriate coding if it is used very frequently in the language as opposed to the frequency of its exceptions. That is the reason why we rely on word frequencies for various ages in our simulations.

Information theory provides a formal version of this assumption: the minimum description length (MDL) principle (Rissanen, 1978). The goal is to minimize the coding cost of both the hypothesis and the data reconstructed from the hypothesis (two-part coding). However, we will see that, in our case, the model contains all the data which lead to a simpler mechanism: the idea is to select the hypothesis which represents the data in the most compact way, that is which has the shortest code length. Given a realization x of a random variable X with probability distribution p, x can be optimally coded with a size of $-\log_2(p(x))$ bits.

For instance, suppose you are exposed to only 4 words A, B, C and D with frequencies .5, .25, .125, .125. For example, exposure could be: BAACADBABACADBAA. An optimal coding would need only 1 bit ($-\log_2(.5)$) to code word A since it occurs 50% of the time. For instance, A would be `0` and all other words would begin with `1`. B needs 2 bits ($-\log_2(.25)$), for instance `10`. C and D both needs 3 bits ($-\log_2(.125)$), for instance `110` for C and `111` for D.

The average code length for a realization of the random variable X is computed by weighting each code length by the corresponding probability. It is exactly what is called entropy:

$$H(X) = -\sum p(x).\log_2(p(x))$$

In the previous example, the average code length is $1\times.5+2\times.25+3\times.125+3\times.125=1.75$ bits

From this point of view, learning is data compression (Grünwald, 2005). To sum up, the general idea of our approach is to generate rules that are coherent with the data observed so far and to select the one with the smallest entropy.

3 Model

Some computational models have been proposed in the literature, but they are concerned with the problem of gender assignment given an existing lexicon rather than dynamically modeling the acquisition of gender knowledge. Their input is therefore a set of words representative of all the words in the language. Analogical modeling (Skousen, 2003) is such a model. It predicts the gender of a new word by constructing a set of words that are analogous to it, with respect to

morphology. Matthews (2005) compared analogical modeling and a neural net and could not find any significant difference. Our model takes noun-gender pairs as input and dynamically updates the set of rules it has constructed so far in order to minimize their description length.

3.1 Input

The input to our model is supposed to represent the noun/gender pairs children are exposed to. We used *Manulex* (Lété et al., 2004), a French lexical database which contains word frequencies of 48,900 lexical forms from the analysis of 54 textbooks. Word frequencies are provided for 3 levels: grades 1, 2 and 3-5.

We used the phonetic form of words[2] because the development of the gender knowledge is only based on phonological data during the first six years of life. It would also be interesting to study the development of written-specific rules, but this will be done in a future work.

We constructed a learning corpus by randomly selecting in this database 200,000 words and their gender such that their distribution is akin to their frequency distribution in *Manulex*. In other words, the probability of picking a given word in the corpus is just its frequency. In fact, we suppose that the construction of the rule depends on the frequency of words children are exposed to and not just on the words at a type level.

It would have been more accurate to take real corpora as input, in particular because the order in which words are considered probably plays a role, but such French corpora for specific ages, large enough to be sufficiently accurate, do not exist to our knowledge.

We now present how our model handles these noun-gender pairs, one after the other.

3.2 Knowledge Representation

Gender knowledge is represented as rules containing exceptions. The premise of a rule is a word ending and the conclusion is a gender. The * character indicates any substring preceding the word ending. A natural language example of a rule is:

(4) */yR/ are feminine nouns (f) except /azyR/, /myR/, /myRmyR/ which are masculine (m).

Exceptions may contain words that could also be organized in rules, which itselves may contain exceptions. Here is an example:

(5) */R/→m except:
/tiRliR/, /istwaR/→f
*/jER/→f except /gRyjER/→m
*/yR/→f except /azyR/ and /myR/→m

The gender knowledge corresponding to a given corpus is represented as a set of such rules. Such a set contains about 80 rules for a grade-1 learning corpus. We now present how this knowledge is updated according to a new noun-gender pair to be processed.

3.3 Rule Construction

Each time a new noun-gender pair is processed, all possible set of rules that are coherent with the data are generated, and the best one, with respect to the minimum description length criterion, will be selected. As an example, consider this little current set of two rules which was constructed from the words /azyR/, /baRaZ/, /etaZ/, /imaZ/, /plaZ/, /SosyR/ and /vwAtyR/[3] (words above below square brackets are the examples which were used to form the rule):

(6) */yR/→f [/SosyR/, /vwAtyR/] except /azyR/→m

(7a) */aZ/→f [/imaZ/, /plaZ/] except /etaZ/, /baRaZ/→m

Then a new word is processed: /kuRaZ/ which is of masculine gender. Since it is not coherent with the most specific rule (rule 7a) matching its ending (genders are different), the algorithm attempts to generalize it with the first-level exceptions in order to make a new rule. /etaZ/ is taken first. It can be generalized with the new word /kuRaZ/ to form the new rule:

(8a) */aZ/→m [/etaZ/, /kuRaZ/]

All other exceptions which could be included are added. The new rule becomes:

(8b) */aZ/→m [/baRaZ/, /etaZ/, /kuRaZ/]

Once a new rule has been created, the algorithm needs to maintain the coherence of the base. It checks whether this new rule is in conflict with other rules with a *different gender*. This is the

[2] We used an ASCII version of the International Phonetic Alphabet.

[3] Translations: /azyR/ (*azur* [azure]), /baRaZ/ (*barrage* [weir]), /etaZ/ (*étage* [floor]), /imaZ/ (*image* [image]), /plaZ/ (*plage* [beach]), /SosyR/ (*chaussure* [shoe]) and /vwAtyR/ (*voiture* [car])

case since we have the exact same rule but for the feminine gender (rule 7a). Conflicting examples are therefore removed from the old rule and put as exceptions to the new rule. In that case of identity between old and new rule, all examples are removed and the rule disappears. The new rule is:

(8c) */aZ/→m [/baRaZ/, /etaZ/, /kuRaZ/] except /imaZ/, /plaZ/→f

After having checked for rules with a *different gender*, the algorithm now checks for existing rules with the *same gender* that the new rule, either more specific or more general. This is not the case here. We thus created our first candidate set of rules (rules 6 and 8c):

CANDIDATE SET #1:
*/yR/→f [/SosyR/, /vwAtyR/] except /azyR/→m
*/aZ/→m [/baRaZ, /etaZ/, /kuRaZ/] except /imaZ/, /plaZ/→f

Other rules could have been generated from the set of exceptions of */aZ/→f. The word /etaZ/ was taken first but the algorithm needs to consider all other exceptions. It then takes /baRaZ/ to form the rule:

(9) */RaZ/→m [/baRaZ/, /kuRaZ/]

Note that this is a more specific rule than the previous one: it is based on a 3-letter ending whereas /etaZ/ and /kuRaZ/ generated a 2-letter ending. No other exceptions can be added. The algorithm now checks for conflicting rules with the same gender and puts this new rule as an exception of the previous rule. Then it checks for possible conflict with rules of different gender, but there are none. The second candidate set is therefore:

CANDIDATE SET #2:
*/yR/→f [/SosyR/, /vwAtyR/] except /azyR/→m
*/aZ/→f [/imaZ/, /plaZ/] except /etaZ/→m
*/RaZ/ [/baRaZ/, /kuRaZ/]→m

Something else needs to be done: removing words from a rule and putting them as exceptions may lead to new generalizations between them or with other existing words. In our case, the algorithm memorized the fact that /imaZ/ and /plaZ/ have been put as exceptions.

It now applies the same mechanism as before: adding those words to the new set of rules, as if they were new words. By the same previous algorithm, it gives the new rule:

(7b) */aZ/→f [/imaZ/, /plaZ/]

In order to maintain the coherence of the rule base, examples of conflicting rules are removed and put as exceptions:

(7c) */aZ/→f [/imaZ/, /plaZ/] except /baRaZ/, /etaZ/, /kuRaZ/→m

We now have our third candidate set of rules:

CANDIDATE SET #3:
*/yR/→f [/SosyR/, /vwAtyR/] except /azyR/→m
*/aZ→f/ [imaZ,plaZ] except /etaZ/, /baRaZ/, /kuRaZ/→m

Figure 1 summarizes the model's architecture.

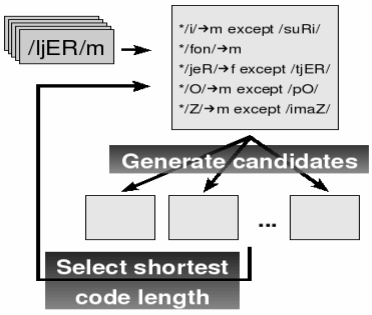

Figure 1. Overall architecture

3.4 Model Selection

This section describes how to choose between candidate models. As we mentioned before, the idea is to select the most compact model. For each exception, we compute its frequency F from the number of times it appeared so far. For each rule, F is just the sum of the frequencies of all examples it covered.

The description length of each rule or exception is $-\log_2(F)$. Since the overall value needs to take into account the variation of frequency of each rule or exception, each description length is weighted by its frequency, which gives the average description length of a candidate set of rules (corresponding to the entropy):

$$weigth(Model) = -\sum F_i.\log_2(F_i)$$

Suppose the words of the previous example were given in that order: /imaZ/ - /vwAtyR/ - /SosyR/ - /imaZ/ - /plaZ/ - /SosyR/ - /plaZ/ - /imaZ/ - /etaZ/ - /vwAtyR/ - /baRaZ/ - /azyR/ - /plaZ/ - /imaZ/ - /imaZ/ - /kuRaZ/

Candidate set #2 would then have an average description length of 1.875 bits:

`azyR m`	$-1/16 \times \log_2(1/16) = .25$
`*yR f SosyR,vwAtyR`	$-4/16 \times \log_2(4/16) = .5$
`*RaZ m baRaZ,kuRaZ`	$-2/16 \times \log_2(2/16) = .375$
`etaZ m`	$-1/16 \times \log_2(1/16) = .25$
`*aZ f imaZ,plaZ`	$-8/16 \times \log_2(8/16) = .5$
	Sum = 1.875 bits

In the same way, candidate set #1 would have a value of 2.18 bits. Candidate set #3 would have a value of 2 bits. The best model is therefore model #2 which is the most compact one, according to the word frequencies.

4 Implementation

For computational purposes, the knowledge internal representation is slightly different than the one we use here: rules and exceptions are represented on different lines such that exceptions are written before their corresponding rules and if a rule is more specific than another one, it is written before. For instance, candidate set #2 is written that way:

```
azyR m
*yR f SosyR,vwAtyR
*RaZ m baRaZ,kuRaZ
etaZ m
*aZ f imaZ,plaZ
```

This allows a linear inspection of the rule base in order to predict the gender of a new word: the first rule which matches the new word gives the gender. For instance, if the previous model were selected, it would predict that the word /caZ/ is feminine, the pseudo-word /tapyR/ is feminine and the pseudo-word /piRaZ/ is masculine.

We could have improved the efficiency of the algorithm by organizing words in a prefix tree where the keys would be in the reverse order of words. However, we are not concerned with the efficiency of the model for the moment, but rather its ability to account for human data.

The algorithm is the following ($R_1<R_2$ indicates that R_1 is more specific than R_2. For instance, */tyR/ is more specific than */yR/, which in turn is more specific than */R/).

```
updateModel(word W, rule base B):
if W matches a rule R∈ B then
  if R did not contain W as an example
    add W to the examples of B
  return B
else
  for all exceptions E of B
    if E and W can be generalized
      create the new rule N from them
      include possible other exceptions
    # More general rule of different gender
    if ∃R∈ B/ R<N and gender(R)≠gender(N)
      put examples of N matching R as exceptions
      memorize those exceptions
      if N now contains one example
        put that example as an exception
      if N contains no examples
        remove N
    # More specific rule of different gender
    if ∃R∈ B/ R≥N and gender(R)≠gender(N)
      put examples of R matching N as exceptions
      memorize those exceptions
      if R now contains one example
        put that example as an exception
      if R contains no examples
        remove R
    # Conflicting rule of same gender
    if ∃R∈ B/ N>R and gender(R)=gender(N)
      include R into N
    if ∃R∈ B/ N<R and gender(R)=gender(N)
      include N into R
    Solutions = {B}
    # Run the algorithm with new exceptions
    for all memorized exceptions E
      Solutions=Solutions ∪ updateModel(E,B)
if no generalizations was possible
  Add W to B
  Solutions = {B}
return(Solutions)
```

5 Simulations

We ran this model on two corpora, representing words grade-1 and grade-2 children are exposed to (each 200,000-word long). 76 rules were obtained in running the grade-1 corpus, and 83 rules with the grade-2 corpus.

Endings	Gender	Gender Predictability	Nb Examples	Nb exceptions
*/l/	f	56%	79	62
*/sol/	m	57%	4	3
*/i/	m	57%	74	55
*/R/	m	72%	188	71
*/am/	f	77%	7	2
*/sy/	m	83%	5	1
*/jER/	f	88%	31	4
*/5/	m	97%	91	2
*/fon/	m	100%	5	0
*/sj6/	f	100%	58	0

Table 1. Sample of rules (with endings and predicted gender) constructed from grade-1 corpus.

Some of the rules of the first set are listed in Table I (from grade-1 corpus). For each rule, represented by a word ending, is detailed its predicted gender, the number of words (as types) following the rule, the number of exceptions. Moreover, the "gender predictability" of each rule is computed (third column) as the percentage of words matching the rule over the total number of words with this ending.

The results of the simulations show that the lengths of word endings vary from only one phoneme (e.g., /*l/, /*i/) to three (/*jER/, /*fon/). These rules do not really correspond to the kind of rules linguists would have produced. They usually consider that the appropriate ending to associate to a given gender is the suffix (Riegel et al., 2005). Actually, the nature of the word ending that humans may rely on to predict gender is an open question in psycholinguistics. Do we rely on the suffix, the last morpheme, the last phoneme? The results of our model which did not use any morphological knowledge, suggests another answer: it may only depend on the statistical regularities of word endings in the language and can vary in French from one phoneme to three and these endings are sometimes matching morphological units.

However, it is worth noting that the model has yet some obvious limitations. The first one is that the gender predictability of rules is variable: while some rules are highly predictive (e.g., */sj§/ 100% feminine, */@/ 97% masculine), other are not (e.g., */l/ 56% feminine, */i/ 57% masculine). The second limitation is that the rules found by our model are accounting for a variable amount of examples. For instance, the rule */R/ masculine accounts for 188 examples while */sol/ masculine does only 4. One could wonder what it means from a developmental point of view to create rules that are extracted from very few examples. Do children build such rules? This is far from sure and we shall have to further address these clear limitations.

Another of our research goals was to test to what extent our model could predict human data. To that end, the model's gender assignment performance was compared to children's one.

6 Comparison to Experimental Data

6.1 Experiment

An experiment was conducted to study how and when French native children acquire regularities between words endings and their associated gender. Nine endings were selected, five which are more likely associated to the feminine gender (/ad/, /asj§/, /El/, /ot/, /tyR/) and four to the masculine gender (/aZ/, /m@/, /waR/, /O/). Two lists of 30 pseudo-words were created containing each 15 pseudo-words whose expected gender is masculine (such as "*brido*" or "*rinloir*") and 15 whose expected gender is feminine (such as "*surbelle*" or "*marniture*"). The presentation of each list was counterbalanced across participants.

Participants were 136 children from Grenoble (all French native speakers): 28 children at the end of preschool, 30 children at the beginning of grade 1, 36 children at the end of grade 1 and 42 children at the beginning of grade 2. Each participant was given a list and had to perform a computer-based gender decision task. Each pseudo-word was simultaneously spoken and displayed in the center of the screen when the determiners "*le*" (masculine) and "*la*" (feminine) were displayed at the bottom of the screen. Then children had to press the keyboard key corresponding to their intuition, which was recorded.

		Pre-school	Beg. Grade1	End Grade1	Beg. Grade2
Endings	Gd.	% Exp. Gd.	% Exp. Gd.	% Exp. Gd.	% Exp. Gd.
/ad/	f	45.24	56.67	67.59**	57.14
/asj§/	f	58.33	58.89	70.37**	65.08**
/El/	f	60.71*	62.22*	76.85**	64.29**
/ot/	f	53.57	71.11**	82.41**	72.22**
/tyR/	f	50.00	68.89**	77.78**	68.25**
/aZ/	m	51.19	64.44**	64.81**	61.11**
/m@/	m	60.71*	55.56	57.41	50.00
/O/	m	61.90*	65.56**	80.56**	78.57**
/waR/	m	52.38	62.22*	64.81**	68.25**

Legend: Gd.:Gender; Beg.:Beginning; % Exp. Gd.:% Expected Gender;

* p<.05,**p<.01

Table 2. Gender attribution rate as a function of endings and grade level.

In brief, results are twofold. First, children have acquired some implicit knowledge regarding gender information associated with word ending. As can be seen in Table 2, at the beginning of grade 1, children respond above chance and in the expected direction for the majority of endings (Chi2 test was used to assess statistical significance). At preschool children responded also above chance for three word endings. Second, there is a clear developmental trend since gender attribution increases in the expected direction with grade level and more endings are determined by the older children. The exposure

to written language during the first school year probably reinforces the implicit knowledge developed by children before primary school.

6.2 Human vs. Model Data Comparison

Two types of analyses were drawn in order to compare model and data. Firstly, the gender predictions obtained from the model were correlated to those given by children, regarding the gender of pseudo-words. Secondly, the endings created by the model were compared to those used in the experimental material. Correlations were computed between our model and human data (Table 3) by taking into account the rate of predicted masculine gender, for each pseudo-word.

	Model Grade 1	Model Grade 2
Preschool	0.31	0.33
Beg. Grade 1	0.6	0.64
End Grade 1	0.82	0.86
Beg. Grade 2	0.74	0.77

Table 3. Correlations between model and data.

The highest correlations are obtained for children at the end of grade 1 and at the beginning of grade 2. This result is interesting since the corpora are precisely intended to represent the lexical knowledge corresponding to the school level of these children. Moreover, the correlations obtained with the grade-2 model are higher (though not significantly) than those obtained with the grade-1 model. It thus seems that our model is fairly well suited to account for children's results, at least for the older ones. The low correlations observed with the younger children of our sample cannot be interpreted unambiguously; one could say that children before grade 1 have not built much knowledge regarding gender of word endings but this conclusion contradicts previous results (Meunier et al., 2008) and it remains to be explored by using a corpora appropriated to the lexicon of preschool children.

The endings used by the model to predict the gender of pseudo-words were also compared with the endings used in the experiment. Table 4 presents these endings as well as the rate of masculine gender predicted for the experimental endings by the two models trained with grade-1 and grade-2 lexicons. First, note that the endings used by the models are the same for both grade-1 and grade-2 lexicons. The growth of the lexicon between grade 1 and grade 2 does not modify these rules. Secondly, one can notice that grade-2 model results are more defined than grade-1 results. Third, a very salient result is that model endings are short. For example, the model did not create a rule such */ad/ and rather used the more compact rule */d/ to predict the gender of the pseudo-word /bOSad/.

	Model Grade 1		Model Grade 2	
Endings	Endings	% Gd. Masc	Endings	% Gd. Masc
/ad/	*/d/	0.28	*/d/	0.17
/asj§/	*/sj§/	0	*/sj§/	0
/El/	*/l/	0.44	*/l/	0.32
/ot/	*/t/	0.14	*/t/	0.09
/tyR/	*/yR/	0.09	*/yR/	0.05
/aZ/	*/Z/	0.8	*/Z/	0.91
/m@/	*/@/	0.95	*/@/	0.98
/O/	*/O/	0.93	*/O/	0.96
/waR/	*/R/	0.72	*/R/	0.82

Table 4. Rate for expected masculine gender predicted by our models.

In fact, the majority of the endings used by the model are short, i.e. composed with one phoneme. Very few endings created by the model are morphological units such as suffixes. In fact, the endings /d/ or /R/ are not derivational morphemes, but the endings /sj§/ or /yR/ are suffixes. So the MDL-based model establishes rules that take into account different types of linguistic units from phonemes to morphemes depending of the statistical predictability of each ending type. This result is related to an important concern about the study of the acquisition of grammatical gender: to which unit do children rely on to predict gender? Do they rely on the last phoneme, biphone, morpheme?

7 Do children rely on morphemes?

In grammatical gender acquisition studies, the kind of endings used often mixes up phonological, derivational and even orthographic cues. Several studies used true suffixes (Marchal et al., 2007, Meunier et al., 2008) to ask children to assign gender to pseudo-words. As those studies consistently showed that children from 3 years old onwards assign a gender to those pseudo-words following the excepted suffix gender, the tentative conclusion was to say that children rely on suffixes to assign the gender of new words. This is an appealing interpretation as the development of morphological structure of words is an important aspect of lexical development and some of this knowledge is acquired very early (Casalis et al., 2000; Karmiloff-Smith, 1979).

However, the observations from the MDL-based model strongly question this assumption: the units retained in the model's rules are often shorter than suffixes and the last phoneme seems often as predictive as the suffix itself as it leads to satisfying correlations with children's data.

So, one would conclude that gender knowledge is not attached to morphological units such as suffix but is rather a knowledge associated with the smaller ending segment that best predicts gender. Note however that despite the high correlations observed, the actual gender predictions issued from children's data and those issued from the model are not exactly of the same magnitude and this would suggest that the MDL-based model presented here must still be worked on in order to better describe gender acquisition. For example, the notion of gender predictability would benefit from being computed from token counts instead of type counts.

8 Conclusion

The purpose of this research was to know which kind of gender information may be constructed and used by children, and which cognitive mechanisms may lead to the construction of such rules. To investigate that issue, we constructed a model based on the MDL principle which reveals to be an interesting way to describe the grammatical gender acquisition in French, although we do not claim that children employ such an algorithm. Our model predicts the gender of a new word by sequentially scanning exceptions and rules. This process appears quite similar to the decision lists technique in machine learning (Rivest, 1987) which has already been combined with the MDL principle (Pfahringer, 1997). However, we are not committed to this formalism: we are more interested in the content of the model rather than its knowledge representation. The comparison between model's results and human data opens a way of reflection on the kind of relevant units on which children would rely on. Perhaps it is not a kind of ending in particular that plays a role but different units varying following the principle of parsimony.

References

Casalis, S., Louis-Alexandre, M.-F. (2000). Morphological analysis, phonological analysis and learning to read French. *Reading and Writing, 12*, 303-335.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology*, 52A, 273-302.

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences, 7*(1), 19-22.

Chater, N., & Vitanyi, P. (2007) 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*, 51(3), 135-163.

Feldman, J. (2003). Perceptual Grouping by Selection of a Logically Minimal Model. *International Journal of Computer Vision,* 55(1), 5-25.

Grünwald, P. (2005). Minimum description length tutorial. In P. D. Grünwald, I. J. Myung & M. Pitt (Eds.), *Advances in MDL: Theory and Applications* (pp. 23-80). Cambridge: MIT Press.

Holmes, V.M., & Segui, J. (2004). Sublexical and lexical influences on gender assignment in French. *Journal of Psycholinguistic Research, 33(6)*, 425-457.

Karmiloff-Smith, A. (1979). *A functional approach to child language.* Cambridge University Press.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers, 36*, 156-166.

Marchal, H., Bianco, M., Dessus, P. & Lemaire, B. (2007). The Development of Lexical Knowledge: Toward a Model of the Acquisition of Lexical Gender in French. *Proceedings of the 2nd European Conference on Cognitive Science*, 268-273.

Matthews, C. A. (2005). French gender attribution on the basis of similarity: A comparison between AM and connectionist models. *Journal of Quantitative Linguistics, 12*(2-3), 262-296.

Meunier, F., Seigneuric, A., Spinelli, E. (2008).The morpheme gender effect. *Journal of Memory and Language, 58,* 88-99.

Pfahringer, B. (1997). Compression-Based Pruning of Decision Lists, in *Proceedings of the 9th European Conference on Machine Learning*, 199-212.

Riegel, M., Pellat, J.C., & Rioul, R. (2005). *Grammaire méthodique du français.* Paris: PUF.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14*, 465-471.

Rivest, R.L. (1987). Learning Decision Lists. *Machine Learning* 2,3 (1987), 229-246.

Skousen, R. (2003). *Analogical Modeling: Exemplars, Rules, and Quantum Computing*. Berkeley Linguistics Society.