The Development of Lexical Knowledge: Toward a Model of the Acquisition of Lexical Gender in French

Harmony Marchal, Maryse Bianco, Philippe Dessus, Benoît Lemaire

In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.) (2007), *Proc. 2th European Cognitive Science Conference (EuroCogSci07)* (pp. 268-273). Mahwah: Erlbaum.

The Development of Lexical Knowledge: Toward a Model of the Acquisition of Lexical Gender in French

Harmony Marchal (Harmony.Marchal@upmf-grenoble.fr)
L.S.E., University of Grenoble-2 & IUFM
38040 Grenoble Cedex 9 FRANCE

Maryse Bianco (Maryse.Bianco@upmf-grenoble.fr)
L.S.E., University of Grenoble-2 & IUFM
38040 Grenoble Cedex 9 FRANCE

Philippe Dessus (Philippe.Dessus@upmf-grenoble.fr)
L.S.E., University of Grenoble-2 & IUFM
38040 Grenoble Cedex 9 FRANCE

Benoît Lemaire (Benoit.Lemaire@imag.fr)
Laboratoire TIMC-IMAG (CNRS UMR 5525), Faculté de Médecine
38706 La Tronche Cedex FRANCE

Abstract

This paper attempts to answer a threefold question: from what kind of cues gender can reliably be assigned to French nouns? How the usage of these sublexical cues is acquired by children? What kind of computational models could account for the human data? We conducted an experiment with 73 1st and 2nd grade children to assess their knowledge about gender information associated with suffixes using a gender decision task. Children had to assign gender to pseudo-words whose endings correspond to French suffixes (*-ture*, *-ment*…). Results show that French children implicitly acquire this knowledge, even before learning to read. Simulation results fit experimental data and show that the word ending represented by suffixes is one of the main cues to assign gender to French words.

Introduction

The aim of this paper is to study how lexical information about words (in our case, gender) is acquired by children, and then to propose a computational model of this acquisition. Very broadly, research aimed at modeling gender assignment in adult speakers is twofold. Some models consider that gender is coded as part of the lexical form of each noun (Schriefers & Jescheniak, 1999), while others consider that sublexical and formal properties of nouns (e.g., phonological, morphological, orthographical) can be related to their gender (Lyster, 2006). These two lines of models are not contradictory, however: the way words are processed in order to be assigned to a given gender remains far from clear. Native speakers are able to assign gender to unknown words (i.e., not represented as lemmas), while L2 French speakers have difficulties to perform this task even for real words.

In French, many word-endings reliably predict gender. For instance, 70% of words respect the following rule: vocalic ending for masculine nouns, consonantal ending for feminine ones (Prodeau & Carlo, 2002). This paper attempts to answer a threefold question: from what kind of cues gender can reliably be assigned to French nouns? How the usage of these sublexical cues is acquired by children? What kind of computational models could account for the human data?

The remainder of this paper is as follows. First, we review literature about the cues by which gender can be inferred, as well as about the development of such abilities. Second, we report an experiment designed to explore whether 6 to 7 year-old children rely on suffixes to build knowledge about gender. Third, we present computational models and how well they fit our experimental data.

Morphological Cues in French Grammatical Gender Assignment

Gender assignment in French can be performed using two main 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). Research shows that phonological (Starreveld & La Heij, 2004), orthographic (Terriault, 2006), or both cues (Holmes & Segui, 2004) are used in such a task.

Taft and Meunier (1998, experiment 1) manipulated two factors: the regularity of the word-ending and the noun frequency. They showed effects of regularity and frequency on students' decision time: regular nouns and high-frequency words were more rapidly assigned than irregular and low-frequency ones, and with less errors. However, no interaction between these two factors was found. This finding is along the same line as Monaghan et al. (2005) claim: "[…] distributional information is most useful for high frequency words, and the artificial language learning ex-

periment indicates that phonological information is compensatory for learning of low frequency words" (p. 178).

In sum, there is a large consensus to consider that "word endings" are powerful cues to gender assignment in French, even if the definition of "word ending" is rather vague and variable across authors. We will nevertheless concentrate on these sublexical cues and especially on morphological suffixes.

Morphological Development and Gender Assignment

Gender information attached to sublexical units is for a large part implicit in nature. This can be illustrated by some recent grammarian's conceptions arguing that word endings are of no use in determining nouns' gender in French (see Lyster 2006). Such a misconception rests undoubtedly on the distributional features of the gender information carried on by word endings. This kind of information can easily be captured by an implicit learning mechanism rather than by a rule-based explicit one. Several linguists have however noticed these regularities. In particular, they described gender information attached to derivational French suffixes (for example words ending in *–age* and *–oir* are masculine whereas those ending in *–elle* and *–ture* are feminine, Riegel, Pellat & Rioul, 2005). The question we address here is to study how and when French native children acquire this knowledge in the course of their lexical development. As gender information attached to derivational suffixes is never explicitly taught in the beginning classes of the primary school, our goal is to study how this information is learned in the course of oral language acquisition and what is the impact of written language learning at the beginning of primary school on this development.

Derivational morphology is largely recognized as a powerful tool for vocabulary growth, at least in later and literate lexicon development. Ravid (2004) argues that before formal written language learning, children have few occasions to be confronted to morphosyntactic forms such as derived nouns, which are rare in oral and everyday language. However, research on morphological awareness has shown that 5-year-old children possess morphological knowledge about words (Colé et al., 2004), especially when this awareness is evaluated through implicit tasks. For example, Casalis and Louis-Alexandre (2000) showed that 5-year-old children succeed at 63.8% in matching a picture to a morphological complex word ("*enrouler*", to roll up), with pictures representing "*enrouler*", "*dérouler*", to unroll, "*rouler*", to roll, "*rouleau*", a roll) spoken by the experimenter. Likewise, when first-grade children read a complex morphological word (e.g., "*laitier*", milkman), reading is easier when children were primed with a morphological associate ("*lait*", milk) rather than with an orthographical associate ("*laitue*", lettuce). These results show that children early acquire some knowledge related to the morphological structure of words and that they can use it in linguistic activities. At least when implicit tasks are used, semantic information attached to affixes seems to be available to young children. Results are less clear when more explicit tasks of morphological awareness, such as word segmentation (to segment the word "*pommade*", ointment, in its morphemes "*pomm–*" and "*–ade*") are involved. In such tasks, children's performance does not exceed 50%. We can conclude that young children have developed morphological knowledge about words before formal learning of written language and that they are especially able to use semantic information captured in morphological units.

It is therefore worth to ask what other kind of information–e.g., grammatical gender–is also early learned by children. Few developmental studies focused on gender acquisition. Tucker et al. (1968) carried out a seminal experiment studying the influence of word-ending on gender assignment. Participants (grade 4 to 7 French native speakers pupils) were asked to assign gender to French words and nonwords (endings were derivational suffixes such as *–aie*, *–ée*, *–é*, *–eur*, *–oir*). Results showed that word-endings are reliably correlated to gender assignment.

Karmiloff-Smith (1979) also carried out a study about the development of grammatical gender. She investigated the ability of children (from 3 to 12 years old) to assign a correct definite article to non-words (the "correct" gender being previously mentioned by means of an indefinite article). She showed that three-year-old children were less able to infer the right definite article without the help of phonological cues (e.g., 100% 4-year-old children assigned correct gender to words like "*une plichette*" compared to 78% to more opaque words, like "*une dilare*"). A complementary experiment showed that when confronted to a discordant information concerning gender (i.e., a masculine article followed by a noun with a feminine-marked suffix), children up to 5 privileged phonological cues over the article.

As in adult experiments, the notion of "word ending" is rather vague and often mixes up phonological, derivational and even orthographic cues. In the Karmiloff-Smith study for example, it is not clear if word-endings such as *–ette* or *–are* are derivational or phonological cues. Early morphological development likely lies on phonological cues and derivational morphology more generally may emerge from the convergence of phonological, orthographical and semantic codes (Seidenberg & Gonnermann, 2000). We will not address this complex issue here but we will rather adopt a classical perspective in isolating derivational units and exploring their impact on gender assignment by grade-1 and 2 children.

Experiment

Our purpose is to examine to what extent 6 to 7 year-old children can rely on derivational suffix to assign gender to nouns. In other words, is gender information encoded with word-endings (potentially suffixes) at the beginning of primary school, before formal learning to read and to spell? To answer this question, we used pseudo-words whose endings represented derivational suffixes (replicating Tucker et al. paradigm with younger children). Using pseudo-words enables to disentangle sublexical processing in gender assignment from lexical processing.

We controlled and manipulated two factors that are not taken into account in the studies reviewed above. The first one concerns the frequency of pseudo-suffixed words, which are non-suffixed words sharing the same word-ending

In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.) (2007), *Proc. 2th European Cognitive Science Conference (EuroCogSci07)* (pp. 268-273). Mahwah: Erlbaum.

as the suffixed words (e.g., the ending *–ment* of *aliment*, food, is not a suffix). Indeed, in the studies reviewed (e.g., Taft & Meunier, 1998), the frequency of word-endings is computed irrespectively of whether they are suffixed or non-suffixed. This variable, called "frequency of non-suffixed nouns" may have an effect on gender assignment. We suppose that if gender information is associated with suffixes this information will be facilitated (built earlier and easier to access) when there are fewer confusing exemplars. In other words, a high frequency of non-suffixed nouns would disturb suffix gender assignment.

The second variable manipulated concerns the "gender exception frequency" of suffixes. We took into account the frequency of suffix gender relative to the total number of nouns sharing the same word-ending but differing in gender (e.g., "*jument*", mare, is feminine while *–ment* is a mostly masculine ending). As Taft and Meunier (1998) showed, low-frequency irregular forms take longer to be assigned a gender than low-frequency regular ones. This variable can predict that when no exception is encountered gender information is acquired earlier.

Furthermore, we controlled the gender of suffixes and a third factor was added to the design: the participants' school level (grades 1 and 2).

Method

Participants Participants were 73 pupils from two schools: 31 children in grade 1 (16 boys and 15 girls) and 42 in grade 2 (25 boys and 17 girls). The experiment was conducted at the beginning of the school year. All children were French native speakers and had no speech or hearing deficiencies.

Material Design First, in order to have an idea about the common nouns known by our participants, we used the *Manulex Infra* database (Peereman, Lété & Sprenger-Charolles, in press). This database is derived from *Manulex*, a web-accessible database listing the word-frequency values for 48,886 lexical entries encountered in 54 French elementary school books, concerning grades 1 and 2. This database presents, for each word, several infralexical variables (syllable, grapheme-to-phoneme mappings, bigrams) and lexical variables (lexical neighborhood, homophony and homography). Second, we selected ten suffixes (5 masculine: *–age*, *–ment*, *–oir*, *–o*, *–ot*; 5 feminine: *–ade*, *–ation*, *–elle*, *–otte*, *–ture*) from a list of thirty French nominal suffixes (available on http://www.etudes-litteraires.com).

The non-suffixed nouns variable had 3 conditions: few [0-400 per million words], medium [401–1000] or high [1001–2000] non-suffixed, while the "gender exception frequency" variable had two: no exception and with exceptions. These values as well as the frequency of non-suffixed nouns were computed from two separate databases (*Manulex Infra* Grades 1 and 2) corresponding to the two school levels.

Material and Procedure For each suffix, 6 pseudo-words were built, using *WordGen* (Duyck et al., 2004). This software generates pseudo-words that are likely to resemble real words while controlling several characteristics–taken from *Manulex Infra*–such as their number of letters or syllables, their number of orthographic neighbors, the frequency of their bigrams. Thus, two lists of 30 pseudo-words were elaborated, 3 words per suffix. The pseudo-words created were five to eleven letters long depending on the suffix and were composed of two or three syllables. Each participant was given a list composed of 15 pseudo-words whose expected gender is masculine (3 words*5 masculine suffixes, such as "*brido*" or "*rinloir*") and 15 pseudo-words whose expected gender is feminine (3 words*5 feminine suffixes, such as "*surbelle*", or "*marniture*"). The experiment was conducted in the children's regular classrooms as part of their school day.

Each participant was presented with the following material, through a computer-based interface. First, pupils were introduced with the task of the experiment, which was to listen to different pseudo-words and to determine their gender. Second, after a five-item practice, each pseudo-word was simultaneously spoken and displayed in the center of the screen, when the articles "*le*" (masculine article in French) and "*la*" (feminine article) were displayed at the bottom of the screen (their position was counterbalanced). Participants then had to press on the corresponding key of the keyboard. The order of word presentation was randomized. Answers and reaction times were recorded.

Data Analysis Each participant's answer was coded 1 (when answer is the expected gender) or 0 (otherwise), then an overall score by suffix was calculated for each participant. We had two dependent variables: the reaction time and the score of correct answers. Analyses of variance were carried out on these two dependent variables. Effects on frequency of non-suffixed nouns, gender exception frequency and school level were tested.

Results

We first looked at gender rate attributed by children for each suffix and tested if the distribution between masculine and feminine attribution differed from chance.

Table 1 shows two interesting results. First, 6 to 7 year-old children have acquired some implicit knowledge regarding gender information associated with suffix. Indeed, children responses are compatible with the expected gender of the majority of suffixes. Gender attribution was above chance level, at first grade for six out of the ten suffixes (*–elle*, *–otte*, *–ture*, *–age*, *–o*, *–oir*), and for eight out of ten at second grade (the same as in first grade plus *–ation* and *–ot*). Second, there is a clear developmental trend since gender attribution in the expected direction is stronger at grade 2 and two more suffixes are determined for the older children. The exposure to written language during the first school year probably reinforces the implicit knowledge developed by children before primary school.

We then computed analyses of variance, using participants (F_1) and items (F_2) as random variables. Table 2 shows the mean scores observed for gender exception frequency, frequency of non-suffixed nouns and school level.

In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.) (2007), *Proc. 2th European Cognitive Science Conference (EuroCogSci07)* (pp. 268-273). Mahwah: Erlbaum.

Table 1: Gender attribution rate as a function of suffix, gender exception frequency and grade level.

			Grade 1			Grade 2		
Suff.	Gd.	GEF	%M	%F	χ^2	%M	%F	χ^2
ade	f	Med.	43.0	57.0	2.67	42.9	57.1	2.57
ation	f	Few	41.9	58.1	2.67	34.9	65.1	11.46***
elle	f	Few	37.6	62.4	6*	35.7	64.3	10.29**
otte	f	Few	31.2	68.8	13.5***	27.8	72.2	24.89***
ture	f	Few	32.3	67.7	13.5***	31.7	68.3	16.79***
age	m	High	64.5	35.5	7.04**	61.1	38.9	6.22*
ment	m	Med.	54.8	45.2	0.67	50.0	50.0	0.00
o	m	High	63.4	36.6	7.04**	78.6	21.4	41.14***
oir	m	Few	62.4	37.6	5.04*	68.3	31.7	16.79***
ot	m	Few	55.9	44.1	0.67	72.2	27.8	24.89***

Legend: Gd.: Gender; Suff.: Suffixes; GEF: Gender Exception Frequency; Med: Medium
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 2: Mean scores and standard deviations on reaction times (s) and correct answers (CA).

			Grade 1				Grade 2			
	GEF	FnSN	CA	SD	RT	SD	CA	SD	RT	SD
Grade 1 DB	With	Few	1.65	1.05	4.63	3.37	1.50	0.97	6.99	3.67
		Med	1.71	0.90	3.90	2.30	1.71	1.04	5.86	3.27
		High	1.92	0.84	4.05	2.31	2.10	0.72	5.92	3.50
	No	Few	1.90	0.80	4.05	2.13	2.06	0.70	5.61	3.05
		Med	1.87	0.66	4.05	2.13	1.99	0.59	5.54	2.77
		High	1.85	0.65	3.83	2.14	2.11	0.65	5.36	2.38

Legend: GEF: Gender Exception Frequency; FnSN: Frequency of non-Suffixed Nouns; DB: Database; Med: Medium

Separate analyses of variance were carried out using the two lexical databases. We observed comparable effects with the two databases, but reaching more often significance with the first grade database. We will only describe these effects.

First, the developmental trend observed on raw data is confirmed. Grade 1 pupils answered more rapidly than grade 2 pupils [$F1(1, 71) = 8.97$, $p = .0038$; $F2(1, 4) = 115.10$, $p < .001$], but no effect of the school level was obtained on correct answers scores.

Second, gender exception has a significant effect by participant not by item, on pupils reaction time [$F1(1, 71) = 11.31$, $p = .0012$] and on correct answers scores [$F1(1, 71) = 6.55$, $p = .0126$]. Mean reaction time is shorter and correct answers scores are higher in the "no exception" condition than in the "with exception" condition, as expected.

Third, we observed that a low frequency of non-suffixed nouns increases the mean reaction times [$F1(2, 142) = 7.21$, $p = .001$] and a high level of non-suffixed nouns increases the correct answers scores [$F1(2, 142) = 4.38$, $p = .0143$]. This result is surprising but can be further explained by the interaction between the gender exception factor and the non-suffixed nouns factor that approaches significance in the participant analysis, on reaction time [$F1(2, 142) = 2.88$, $p = .0593$] and on correct answers scores [$F1(2, 142) = 2.82$, $p = .0628$]. It shows that in the "no exception" condition, pupils' reaction time and correct answers scores vary little according to the modalities of the "non-suffixed" variable, while pupils' reaction time is higher and correct answers scores are weaker in the "few non-suffixed nouns" and "with exceptions" conditions.

Computational Cognitive Models of Gender Assignment

We also performed computer simulations to account for these results. Computational models of gender assignment have already been presented and compared to human data in the literature (Eddington, 2002), even for the French language (Matthews, 2005) but, as opposed to our approach, they aim at modelling the adult behavior. Therefore, the input of these models is generally a set of words representative of all words in the language. In our developmental approach, we need to identify the words children are supposed to know in first and second grade. Consequently, we relied on the same database we used for designing the experimental material in order to get an estimate of the words children know in 1st and 2nd grade. We tested three models of growing complexity.

Occurrence-Based Modeling

The first model we built was quite simple. It only relies on suffixes to predict gender. Basically, children would associate a probability of being masculine or feminine to each suffix, according to the words they were exposed to. In this simple model, the probability of a suffix being masculine is just the proportion of masculine words in the set of all words of that suffix the child knows. The probabilities of a given suffix S being masculine or feminine are:

$p_{\text{masculine}}(S) =$ number of masculine words ending in S / number of words ending in S

$p_{\text{feminine}}(S) = 1 - p_{\text{masculine}}(S)$

For each of the suffixes of our experiment, we computed its probability of being masculine. We then calculated correlations with children data. We found interesting significant correlations of .65 for 1st grade and .77 for 2nd grade. This result means that children heavily rely on suffixes to predict gender at least in 2nd grade. This is coherent with the fact that the ending of the word is the best predictor of gender in French (Tucker et al., 1968). However, the ending of a word that children would store in memory might not be the last suffix but rather its last phoneme, its last syllable, etc. Which of these children rely on?

In order to answer that question, we modified the model for taking into account the last phoneme or the last syllable instead. Correlations with children data were not higher: when the model relied on the last syllable, correlations were .64 for 1st grade and .75 for 2nd grade whereas results were .52 for the 1st grade and .63 for the 2nd grade when the model relied on the last phoneme. We need to obtain more

In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.) (2007), *Proc. 2th European Cognitive Science Conference (EuroCogSci07)* (pp. 268-273). Mahwah: Erlbaum.

data to investigate this important issue that we will discuss in the conclusion part.

Frequency-Based Modeling

Word frequency should also play a role in gender assignment. Let us present the case of the suffix *–age*. In French, most of the words ending in *–age* are masculine. There are a few exceptions, but their frequencies are quite high in particular in the children lexicon (*image*, *cage*, etc.). These high frequencies should interfere with the normal assignment of a non-word ending in /age/ as masculine. How could we account for that in order to test that hypothesis? We modified the previous model in order to take into account word frequencies. This second model is more coherent with the analysis of children data in which variables also relied on frequencies. These frequencies were also obtained from the *Manulex* database. We normalized by taking the log of these frequencies and summed up for all words.

$p_{\text{masculine}}(S)$ = sum of frequency of masculine words ending in S / sum of frequency of words ending in S

$p_{\text{feminine}}(S) = 1 - p_{\text{masculine}}(S)$

Correlations with human data were almost the same as in model 1. Although some suffixes got different probabilities, the difference was quite small. For instance, the probability of a word ending in *–age* being masculine only decreased from .94 in model 1 to .87 in that model. It might be the case that word-endings frequency does not play the role we expected it to do. Word frequency has a main role in many cognitive processes, but is it the same with word-ending frequency?

Analogical Modeling

We previously showed the importance of the word ending in the gender assignment. However, it is likely that the rest of the word would somehow play a role. In order to investigate that issue, we implemented and tested a cognitive model of gender assignment that would take into account all parts of the word. The analogical model (AM, (Skousen, 1989) is an exemplar-based model which has been used to predict several psycholinguistic processes, in particular gender assignment (Eddington, 2002). AM is based on the idea that the gender of a new word can be identified from a set of words that are analogous to it. This analogy is drawn by considering each letter (or each phoneme) of the new word as a variable. By considering some of these variables as placeholders (or empty slots), AM progressively generalizes the new word in order to get examples that match it. A generalization is interesting if all examples that it matches belong to the same category. Let us take an example.

Suppose we want to assign a gender to the non-word "*prage*". We first put only one placeholder at a time, which gives us 5 generalizations: /*rage/, /p*age/, /pr*ge/, /pra*e/ and /prag*/. The first one matches with 9 examples that are all masculine (*barrage*, *courage*, etc.). This first generalization is therefore called deterministic; its examples will be kept as analogical examples in what is called an analogical set. The second generalization matches two conflicting examples, one masculine (*péage*) and one feminine (*plage*), which is not good for considering it for analogy. However, AM makes an exception in this case and keeps generalizations that are not deterministic but whose ancestors do not match any example. Once all generalizations with only one placeholder have been considered, AM attempts to generalize a bit more by putting two placeholders to all generalizations that have been successful. For instance, the generalizations /**age/, /*r*ge/, /*ra*e/ and /*rag*/ will be constructed from /*rage/. In our 1st grade database, 63 examples match /**age/, only two of them being feminine. This generalization will however not be kept for further generalization because it is not homogeneous. The next generalization, /*r*ge/, matches 9 examples (*orage*, *naufrage*, etc.) that are all masculine. This generalization is therefore interesting and will be kept for further generalization, leading to /***ge/, /*r**e/ and /*r*g*/. Once no more generalization can be found, the prediction is made according to the analogical examples found. In our case, the probability is the proportion of analogical examples of the corresponding gender:

$p_{\text{masculine}}(S)$ = number of masculine examples analogous to S / number of examples analogous to S

$p_{\text{feminine}}(S) = 1 - p_{\text{masculine}}(S)$

This method is interesting in that examples that seem far from the given word can contribute to the prediction, because what distinguish them from the given word are variables that do not change the gender. This is not the case with a nearest neighbor algorithm which only relies on examples that resembles the given word. We implemented the model in *Perl*, ran it on the 1st and 2nd grade lexicon and obtained the following significant correlations with the children data: .60 for the 1st grade and .66 for the 2nd grade. These results are not better than those obtained with the model 1.

We then realized that the AM algorithm was too rigid from the sake of cognitive plausibility. In the previous example, the generalization /**age/ will be ruled out because the examples it matches (61 masculine and 2 feminine) do not belong to the same gender. However, it is possible that this generalization would be used by children for predicting a masculine word because 97% of the examples are masculine. We then modified the AM algorithm and considered a generalization as homogeneous if more than 95% of the examples it matches belong to the same gender. We also changed the matching algorithm of the initial AM model. The idea is that a word like "*bousculade*" (rush) might serve as an example for predicting the gender of the non-word "*boucade*", which is not the case in the AM algorithm since the matching is done variable per variable. In our new algorithm, a letter of the given word can be replaced by several letters in the examples. Correlations with children data did not change for the 1st grade and were slightly higher for the 2nd grade: .72 but the difference is not significant. Although

In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.) (2007), *Proc. 2th European Cognitive Science Conference (EuroCogSci07)* (pp. 268-273). Mahwah: Erlbaum.

we could not conclude anything from this result, we believe the scientific community would benefit from this idea of a fuzzy matching in the AM model.

Last but not least, we also applied the algorithm on the phonetic form of the data. Matthews (2005) did not obtained better results in applying the AM model on phonetic data, but it was worth trying it on children data. All examples and test items were considered as sequences of phonemes. We obtained correlations of .57 for 1st grade and .74 for the 2nd grade, which is always in the same range of values.

Conclusion

Our experiment showed that French children implicitly acquire knowledge to predict gender, even before learning to read. There was an increase of performance during the 1st grade, while no knowledge about cues was taught at school, possible as an effect of the exposition to written material. Our simulations showed that the word ending represented by suffixes is likely the main information from which gender can be predicted in French: more complicated models, as well as models using other types of units (syllable, phoneme) did not succeed in better accounting for children data. However, the nature of word ending remains to be more precisely defined. For example, one might ask why children would rely on predefined endings like syllables, bigrams or morphemes as we researchers would expect. They could instead use the most efficient ending depending on its prediction power. If the last letter or the last phoneme is enough to predict gender, then why would we store more information? However, in some cases, a longer string of letters is necessary to predict gender. For instance, the ending *–on* is obviously not enough to predict the gender since, in our 1st-grade data, there are 116 feminine and 131 masculine nouns with the same ending. The trigram /ion/ is much better: 98 feminine and 9 masculine nouns. /tion/ is really a good unit to code: no masculine nouns at all. Therefore, why would we store a specific information for /ation/?

Acknowledgments

We are grateful to Laurent Lima for his statistical help. We also would like to thank teachers from Barraux and Domène who kindly allowed us to experiment in their classrooms.

References

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

Colé, P., Royer, C., Leuwers, C. & Casalis, S. (2004). Les connaissances morphologiques dérivationnelles et l'apprentissage de la lecture chez l'apprenti-lecteur du C.P. au C.E.2. *L'Année Psychologique, 104,* 701-750.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). A tool for word selection and non-word generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments & Computers*, *36*(3), 488-499.

Eddington, D. (2002). Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics, 9*, 49-75.

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: Cambridge University Press.

Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *French Language Studies*, 16, 69-92.

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.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*, 143-182.

Peereman, R., Lété, B., & Sprenger-Charolles, L. (in press). Manulex-Infra: Distributional characteristics of grapheme-phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods, Instruments and Computers*.

Prodeau, M., & Carlo, C. (2002). Le genre et le nombre dans des tâches verbales complexes en français L2: grammaire et discours. *Marges Linguistiques*, *4*, 165-174.

Ravid, D. (2004). Later lexical development in Hebrew: derivational morphology revisited. In R.A. Berman (Ed.), *Language development across childhood and adolescence* (pp. 53-82). Amsterdam: John Benjamins.

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

Schriefers, H., & Jescheniak, J. D. (1999). Representations and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research*, *28*(6), 575-599.

Seidenberg, M.S. & Gonnermann, L.M. (2000). Explaininng derivational morphology as the convergence of codes. *Trends in Cognitive Science, 4*(9), 353-361.

Skousen, R. (1989). *Analogical Modeling of Language*. Dordrecht: Kluwer.

Starreveld, P. A. & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes*, *19*(6), 677-711.

Taft, M., & Meunier, F. (1998). Lexical representation of gender: A quasiregular domain. *Journal of Psycholinguistic Research*, *27*(1), 23-45.

Terriault, L. (2006). L'attribution passive et l'acquisition du genre grammatical en français langue seconde. *Proc. Conf. CeSLa'06.* Montreal.

Tucker, G. R., Lambert, W. E., Rigault, A., & Segalowitz, N. (1968). A psychological investigation of French speakers' skill with grammatical gender. *Journal of Verbal Learning and Verbal Behavior*, *7*(2), 312-316.