French Language Studies **16** (2006), 69–92, © Cambridge University Press
doi:10.1017/S0959269506002304

Predictability in French gender attribution: A corpus analysis[1]

ROY LYSTER
McGill University

(Received October 2004; revised November 2005)

ABSTRACT

This article presents a corpus analysis designed to determine the extent to which noun endings in French are reliable predictors of grammatical gender. A corpus of 9,961 nouns appearing in *Le Robert Junior Illustré* was analysed according to noun endings, which were operationalised as orthographic representations of rhymes, which consist of either a vowel sound (i.e., a nucleus) in the case of vocalic endings or a vowel-plus-consonant blend (i.e., a nucleus and a coda) in the case of consonantal endings. The analysis classified noun endings as reliably masculine, reliably feminine, or ambiguous, by considering as reliable predictors of grammatical gender any noun ending that predicts the gender of least 90 per cent of all nouns in the corpus with that ending. Results reveal that 81 per cent of all feminine nouns and 80 per cent of all masculine nouns in the corpus are rule governed, having endings that systematically predict their gender. These findings, at odds with traditional grammars, are discussed in terms of their pedagogical implications.

1 INTRODUCTION

More than thirty-five years ago, Tucker, Lambert, Rigault and Segalowitz (1968: 316) wrote that 'French grammarians have been hasty in their conclusion that there are no regularities or only minimal ones to gender determination'. Years later, however, French grammarians continue to uphold the unsystematic nature of gender attribution. For example, Bérard and Lavenne (1991: 18) write that 'criteria based on word endings (e.g., vowel vs consonant) will be of no use for knowing

[1] This research was supported by grants from the Social Sciences and Humanities Research Council of Canada (nos. 410-98-0175 and 410-2002-0988). A version of this study was presented at the annual meeting of the American Association for Applied Linguistics in Portland, Oregon, on May 3, 2004. I gratefully acknowledge Susan Ballinger and Michel Gagnon for their contributions as Research Assistants, and Walcir Cardoso, Murray Munro, Hubert Séguin and *JFLS* reviewers for helpful comments on earlier versions. Correspondence concerning this article should be emailed to the author at roy.lyster@mcgill.ca.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

whether a word is masculine or feminine'.[2] Similarly, Jacob and Laurin (1994: 145) report that 'the gender of a noun referring to an inanimate object follows no strict rule'[3] and Bosquart (1998: 112) advises that 'no rule actually governs the gender of names of things or concepts'.[4] Learners of French as a second or foreign language (L2) are thus faced with the challenge of sorting out what is claimed to be an arbitrary subsystem of French grammar, left to their own devices to learn gender attribution of inanimate nouns on an item-by-item basis.

Contrary to assertions about the arbitrary nature of grammatical gender, Tucker, Lambert and Rigault (1977) reported that grammatical gender entails a highly rule-governed subsystem, in which 'distinctive characteristics of a noun's ending and its grammatical gender are systematically related' (Tucker *et al.*, 1977: 64). Their report and others (Tucker, 1967; Rigault, 1971) resulted from lexicographical analyses of 26,725 nouns in the *Petit Larousse* made possible by Quemada's (1971) inverse dictionary listing all 31,619 nouns in the *Petit Larousse* congregated by written ending and separated according to gender.

Motivated by questions left unanswered by analyses of Quemada's *Petit Larousse* corpus, the present study analyses a smaller corpus, namely, *Le Robert Junior Illustré* (1994, 1999), a 20,000-word dictionary designed for 8–12-year-old children and henceforth referred to as RJ. Specifically, the present study aims to determine the overall reliability of noun endings as predictors of grammatical gender in a given corpus and, at the same time, to ascertain the proportion of nouns with endings unable to predict gender in the same corpus. More generally, the present study was undertaken to explore results yielded by analyses of a corpus that is less extensive than the *Petit Larousse* corpus and thus more appropriate for L2 learners: namely, a dictionary such as RJ with fewer entries for nouns considered rare, archaic, too technical, overly erudite, or simply obscure. In addition to being considered an appropriate corpus of reasonable size and with potential for providing useful information about gender to L2 learners, RJ was selected because its CD-ROM version enabled computerised searches of nouns by their endings and gender. Other sources of relevant information were also consulted (e.g., Surridge, 1995; Huot, 1991; see also Batchelor and Offord, 1993; Hawkins and Towell, 2001; Ollivier, 1986), but none reported full quantitative analyses to reveal the overall predictive value of noun endings in French.

The prospect of noun endings as indicators of gender is not without some debate. In addition to French grammarians who maintain that gender attribution of inanimate nouns is, for the most part, arbitrary and unsystematic, researchers are not in agreement as to the nature of gender attribution rules, if rules indeed exist. Whereas Tucker *et al.* (1977) referred to noun endings as important indicators of grammatical gender, Carroll (1989: 563) claimed that 'Endings do

[2] 'Des critères portant sur le final des mots (du type voyelle/consonne) ne vous seront d'aucune utilité pour savoir si un mot est masculin ou féminin'.
[3] 'Le genre du nom désignant un objet inanimé ne suit pas de règle stricte'.
[4] 'Aucune règle ne préside en effet au genre des noms de choses ou de concepts'.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

not exist'. She suggested that Tucker *et al.* (1977) confound phonology and morphology, and argued that the rules are more accurately characterised as morphological, being sensitive to suffixes of complex nouns. Surridge (1995) also argued that morphological rules are more reliable than phonological rules, but limited phonological rules exclusively to final phonemes and morphological rules to compounds and clearly segmented suffixes such as *-scope* and *-phone* (see also Müller, 1990). Karmiloff-Smith (1979: 167) referred interchangeably to noun suffixes and noun endings as reliable clues, and concluded that young French-speaking children (3–4 years old) develop 'a very powerful, implicit system of phonological rules, based on the consistency . . . of phonological changes in word endings' (see also Möhring, 2001), thus pointing to an interaction between phonology and morphology, captured as well in Harley's (1998) reference to morphophonological rules of gender attribution.

Much of the debate thus centres on whether gender attribution is derived phonologically or morphologically, and also concerns how native speakers of French actually process gender in psycholinguistic terms. Carroll (1989) argued that native speakers of French acquire gender attribution as an inherent property of nouns by initially processing nouns and their gender-specific determiners as co-indexed chunks, whereas anglophone learners of French L2, influenced in part by their first language (L1), process nouns and their determiners as separate lexical items. Tucker *et al.* (1977) described a 'backward processing' technique whereby native speakers focus on the final phoneme then process back (from right to left) until a clear decision about gender attribution can be made. They reported on a series of experiments demonstrating that native speakers' skill in assigning gender to both attested and nonce forms was an example of rule-governed behaviour in that their choices invariably reflected the distributional regularities between gender attribution and noun endings, as observed in the *Petit Larousse* analyses. Desrochers, Paivio and Desrochers (1989) found that native speakers retrieve gender more quickly for high-frequency nouns with reliable endings than for low-frequency nouns with less reliable endings, suggesting that noun endings as indicators of gender attribution are processed in relation to their frequency. Holmes and Dejean de la Bâtie (1999: 498) confirmed this finding, but concluded that 'knowledge of word endings is not the primary basis for gender attribution', precisely because native speakers were observed to be slower and less accurate in assigning gender to non-words than to real words. They argued instead that other gender-identifying information, such as that found in determiners, is accessed first.

Irrespective of different views to explain native speakers' facility with gender attribution, most researchers agree that L2 learners of French can benefit from information about word-internal properties that affect grammatical gender (Carroll, 1989; Harley, 1979, 1998; Holmes and Dejean de la Bâtie; Lyster, 2004b; Tucker *et al.*, 1977; Séguin, 1969; Stevens, 1984; Valdman, 1976). Researchers agree that learning grammatical gender is an extreme example of the difference between L1 and L2 acquisition, as attested by the seemingly effortless and flawless acquisition

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

of grammatical gender by native speakers of French on the one hand, and, on the other, the notoriously difficult and often incomplete acquisition of this same grammatical subsystem by many L2 learners of French. Tucker *et al.* (1977: 11) remarked that 'the necessity to master grammatical gender may be the single most frustrating and difficult part of the study of French as a second language'.

The present study has as its goal to provide learners and teachers of French L2 with helpful information about gender attribution, derived from a corpus-based analysis, without regard for how native speakers of French actually acquire and process gender attribution. The specific research questions are formulated as follows:

1. Which noun endings in the RJ corpus predict grammatical gender and to what extent?
2. What proportion of nouns in the RJ corpus has endings that reliably predict grammatical gender?

2 PROCEDURE

This corpus-based study was designed to determine the extent to which noun endings in French are reliable predictors of grammatical gender. The corpus derives from the CD-ROM version of *Le Robert Junior Illustré* (1999), a 20,000-word dictionary designed for 8–12-year-old children, which enabled computerised searches of nouns by their endings and gender.

2.1 The corpus

The RJ corpus used for analysis comprises 9,961 nouns, including singular nouns only and excluding any nouns that occur strictly in plural forms (e.g., *les moeurs, les environs, les confins*). Proper names were also excluded, as were nouns representing cardinal and ordinal numbers (e.g., *un dix, un quarante, un quarantième*). Compound nouns such as verb-plus-noun compounds (e.g., *un casse-tête*), which are always masculine regardless of noun ending (Surridge, 1985, 1986), were excluded as well. Included, however, were adjective-plus-noun compounds (e.g., *une longue-vue, un beau-père*), because their gender attribution is normally determined by the noun, which occurs in final position. Because the overall purpose of the corpus analysis was to describe rule-governed patterns, nouns with endings occurring only once in the corpus were also excluded, reducing the original corpus of 10,112 nouns to 9,961 nouns, through exclusion of high-frequency nouns such as *le coude, le doigt, la jambe, le nid, la paix, le riz,* and others such as *l'algue, la bravoure, le glaive, le hareng, le luth, le pouls, la valse* and *le zinc.*

Initially, only inanimate nouns were to be targeted in this corpus analysis, based on the premise that gender attribution in the case of animate nouns is relatively straightforward and also well documented. However, the corpus was soon expanded to include both animate and inanimate nouns, because it became clear that many animate nouns conform to the same structural patterns under investigation, both in cases where a distinction between sexes is made (e.g., *un chien, une chienne; une*

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

vache, un taureau) and in epicene forms where no such distinction is made (e.g., *une coccinelle, une baleine, un saumon*).

The North-American hard copy version of *Le Robert Junior Illustré* (1994) was used as a secondary source to facilitate use of the resulting word bank in local schools. Use of this secondary source reduced the number of anglicisms otherwise found in the European version (e.g., *le bowling, un gentleman, un meeting, le standing, une star, le week-end, un zoom*) and added various nouns such as those depicting flora (e.g., *une épinette, une pruche*) and fauna (e.g., *une barbotte, un béluga, une outarde*), many of which derive from Amerindian languages (e.g., *un achigan, un atoca, un cacaoui, un maskinongé, un ouaouaron*), as well as various cultural phenomena associated with, for example, weather (*la poudrerie*), education (*un cégep*), clothing (*une tuque*) and food (*une tourtière*). In addition, the North-American version includes a wider range of feminine forms designating professions (e.g., *une camionneuse, une colonnelle, une ingénieure, une pompière, une professeure*).

2.2 The analysis

The RJ corpus comprises 9,961 nouns organised according to the predictive value of their endings. Noun endings in this analysis refer to variable orthographic representations of rhymes, defined as either a vowel sound (i.e., a nucleus) in the case of vocalic endings, or a vowel-plus-consonant blend (i.e., a nucleus and a coda) in the case of consonantal endings.

Predictive value, defined here as the extent to which noun endings accurately predict gender attribution, is calculated as a percentage of the number of nouns in which a set of endings accurately predicts gender attribution, relative to the total number of nouns in the corpus with the given endings:

$$\text{Predictive value} = \frac{\text{No. of nouns with a given set of endings that predict their gender}}{\text{No. of nouns with the given endings}} \times 100$$

The analysis classifies noun endings as follows:

- Feminine endings predict feminine gender for at least 90 per cent of nouns with these endings;
- Masculine endings predict masculine gender for at least 90 per cent of nouns with these endings;
- Ambiguous endings indicate that fewer than 90 per cent of nouns with these endings are of only one gender.

A reliable predictor of grammatical gender is therefore any noun ending that predicts the gender of least 90 per cent of all nouns in the corpus with that ending. The lower limit was set at 90 per cent with the expectation that such highly predictive rules would be of considerable use to learners and teachers of French L2.

All nouns in the corpus have been classified as either feminine (F), masculine (M), or both feminine and masculine (F/M). Nouns with dual gender include a small set of inanimate nouns, more accurately described as homophones which

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

happen to differ in gender (Valdman, 1976), whose meanings change according to gender (e.g., *crêpe, livre, manche, mémoire, poste, somme, tour*), and a larger set of animate nouns with epicene forms (e.g., *athlète, élève, garde, gastronome, rebelle, touriste*), which refer to both sexes and adopt one gender or the other to do so, but without changing orthographic form as do non-epicene forms (e.g., *copain/copine*, *ami/amie*).

3 RESULTS

Final phonemes are presented in Table 1 in decreasing order of their predictive value, ranging from 99 per cent for /ã/ to 53 per cent for /e/. Six final phonemes, accounting for only 18 per cent of the entire corpus, are reliable predictors of grammatical gender with respect to the 90 per cent criterion: /ã/, /ɛ̃/, /o/, and /ɛ/ reliably predict masculine gender, while /z/ and /ʃ/ reliably predict feminine gender. Other noteworthy predictors falling in the 78–88 per cent range are /ø/, /u/, /ʒ/, /a/, /f/, and /m/ for masculine gender, and /d/, /n/, /s/, /t/, and /v/ for feminine gender. Noteworthy is that phonemic predictors of masculine gender include a majority of vocalic phonemes, while phonemic predictors of feminine gender include only consonantal phonemes. As previously stated, however, there are word-internal patterns that predict grammatical gender more reliably than final phonemes alone. That is, final phonemes on their own are not consistently the most reliable predictors; instead, they interact with orthographic representations, which surpass phonological representations in terms of predictive value. The remainder of this section is devoted to identifying orthographic representations that are predictive of gender attribution. Final phonemes are used as an organizational aid throughout this section as a convenient way of presenting the results, beginning with vocalic final phonemes followed by consonantal final phonemes, grouped according to whether or not the phonemes are predominantly associated with one gender or another. A predominant association means that, in the case of vocalic final phonemes, more than 85 per cent of nouns ending with that phoneme are of one gender (Table 2) and, in the case of consonantal final phonemes, more than 65 per cent of nouns ending with that phoneme are of a particular gender (Tables 4 and 5). The ordering of final phonemes within each table reflects, in decreasing order, the number of nouns in the corpus ending with the given phoneme.

Throughout the tables and related text, C refers to any consonant and V to any vowel. That diphthongal vowels function as monophthongs is taken into account in cases where predictive value is affected. For example, the vocalic ending *-ion* and the consonantal ending *-ière* are each classified as rhymes, with /j/ functioning as a semivowel (but as a consonant in noun-final position as in *fille*). This allows important distinctions to be made between diphthongal and monophthongal nuclei that affect gender attribution: for example, *-ion* vs *-on* and *-ière* vs *-ère*.

Predictive values displayed in the remaining tables are calculated on the basis of nouns having only one gender, never two, and are thus on the conservative side. That is, predictive values would be higher if nouns with dual gender were counted

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 1. *Final phonemes in descending order of their predictive value in RJ corpus*

Final phoneme	Number of nouns ending with phoneme	Per cent of nouns whose gender is predicted by final phoneme
/ɑ̃/	675	99 per cent M
/ɛ̃/	243	98 per cent M
/z/	239	97 per cent F
/o/	312	93 per cent M
/ɛ/	239	93 per cent M
/ʃ/	105	90 per cent F
/ø/	24	88 per cent M
/u/	68	87 per cent M
/ʒ/	303	87 per cent M
/d/	230	86 per cent F
/a/	259	85 per cent M
/f/	71	82 per cent M
/n/	348	82 per cent F
/m/	249	80 per cent M
/s/	598	79 per cent F
/t/	679	79 per cent F
/v/	68	78 per cent F
/ɔ̃/	1,061	71 per cent F
/i/	523	68 per cent F
/j/	143	65 per cent F
/p/	66	64 per cent F
/y/	95	63 per cent M
/R/	1,507	63 per cent M
/g/	54	61 per cent F
/k/	181	56 per cent M
/ŋ//ɲ/	33	55 per cent F
/l/	561	54 per cent M
/b/	26	54 per cent F
/e/	1,001	53 per cent M
Total	9,961	

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 2. *Orthographic endings with vocalic final phonemes that are predominantly masculine*

Final phoneme	Orthographic endings	Number of nouns by gender				Predictive value
		F	M	F/M	Total	
/ã/	Masc (*volcan*): *-an, -anc, -and, -ang, -ant, -ent, -amp, -emps, -aon*	4	670	1	675	99 per cent
/o/	Masc (*rideau*): *–eau, -au, -aud, -aut, -o, -oc, -op, -os, -ot, -ow*	15	288	6	309	93 per cent
	Amb (*taux/chaux*): *-aux*	2	1	0	3	–
/ɛ̃/	Masc (*patin*): *-in, -ing, -int, -aim, -um/-un/-unt*	4	239	0	243	98 per cent
/ɛ/	Fem (*haie*): *-aie*	16	0	0	16	100 per cent
	Masc (*délai*): *-ai, -ais, -ait, -ès, -et, -ect, -ey*	0	221	0	221	100 per cent
	Amb (*forêt/genêt*): *-êt*	1	1	0	2	–
/u/	Fem (*boue*): *-oue*	8	0	0	8	100 per cent
	Masc (*bijou*): *-ou, -oût, -out, -oup, -oux*	1	59	0	60	98 per cent
/Ø/	Fem (*banlieue*): *-eue*	3	0	0	3	100 per cent
	Masc (*feu*): *-eu/-œu, -eud/-eux*	0	21	0	21	100 per cent

as being of one gender at least half the time, as was the case for final phonemes in Table 1.

3.1 Vocalic final phonemes

Table 2 identifies orthographic endings with vocalic final phonemes that are predominantly masculine. Of these, nouns ending with the phoneme /ã/ are the most frequent ($n = 675$) and most predictive: 99 per cent of nouns with these endings (e.g., *-an, -ent*) are masculine. Nouns ending with the phoneme /ɛ̃/ are less frequent ($n = 243$) but similarly predictive: 98 per cent of nouns with these endings (e.g., *-in*) are masculine. Orthographic representations of the final phoneme /o/ that are considered masculine (e.g., *-eau, -o*) predict masculine gender for 93 per cent of the 309 nouns with these endings. Noteworthy is that the small number of feminine nouns ending in *-o* (e.g., *une photo, une moto*) are mostly truncated forms derived from rule-governed feminine nouns (*une photographie, une motocyclette*). The final phonemes /ɛ/, /u/, and /Ø/ are also strong predictors of masculine gender. All 221 nouns with masculine endings representing /ɛ/ (e.g. *-ais, -et*) are masculine; 59 of 60 nouns with masculine endings representing /u/ (e.g., *-ou, -out*) are masculine (the

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 3. *Orthographic endings with vocalic final phonemes that are feminine, masculine, or ambiguous*

Final phoneme	Orthographic endings	Number of nouns by gender				Predictive value
		F	M	F/M	Total	
/ɔ̃/	Fem (*station*): *-ion*	721	15	0	736	98 per cent
	Masc (*pont*): *-om, -omb, -onc, -ond, -ong, -ont, -on* (but not after *s/ç*)	2	255	0	257	99 per cent
	Amb (*glaçon/leçon*): *-çon, -son*	33	35	0	68	–
/e/	Fem (*qualité*): *-té, -ée, -ié*	465	24	2	491	95 per cent
	Masc (*plancher*): *-er, -ed, -é* after C (C≠*t*)	2	495	0	497	99.6 per cent
	Amb (*pitié/allié*): *-ié*	4	9	0	13	–
/i/	Fem (*scie*): *-ie*	352	6	0	358	98 per cent
	Masc (*défi*)*: -i, -il, -it, -is, -ix, -y(s)*	4	160	1	165	97 per cent
/a/	Fem (*soie*): *-oie*	8	1	0	9	89 per cent
	Masc (*drap, bois*): *-ap, -at, -ac, -as, -ois, -oit, -oid(s)*	1	134	0	135	99 per cent
	Amb (*caméra/opéra*): *-a, -ia, -oi, -oix*	30	83	2	115	–
/y/	Fem (*rue*): *-ue*	31	0	0	31	100 per cent
	Masc (*tissu*): *-u, -us, -ut, -ux*	4	60	0	64	94 per cent

only exception is *la toux*); and all 21 nouns with masculine endings representing /Ø/ (e.g., *-eu*) are masculine. These three phonemes illustrate clearly that orthography overrides phonemic ending in predicting gender: all nouns ending in *-aie*, *-eue*, and *-oue* are feminine, whereas all nouns ending in *-ai*, *-eu*, and *-ou* are masculine.

Table 3 presents orthographic endings with vocalic final phonemes that are not identified with only one gender. The final phonemes /ɔ̃/ and /e/ deserve special attention, first, because nouns with these endings together comprise more than one-fifth of the entire corpus and, second, because their analysis diverges somewhat from that of other final vocalic phonemes in that, in two cases, the predictive value of orthographic representations of rhymes (viz. *-on* and *-é*) is relatively low and becomes reliable only when syllabic onset is considered.

In the case of /ɔ̃/, of 31 nouns whose final phoneme is represented orthographically by *-om* (e.g., *un prénom*), *-omb* (e.g., *le plomb*), or *-on* plus a mute consonant (e.g., *un pont*), 100 per cent are masculine. Of 736 nouns ending with the diphthongal nucleus *-ion*, 98 per cent are feminine (e.g., *une station*); the 15 exceptions include nouns such as *bastion*, *lampion* and *camion*. If we consider more than only rhymes and include the whole syllable as the final segment, the predictive value of nouns ending in *-sion* reaches 100 per cent. Similarly, the orthographic ending *-on* also increases its predictive value if the preceding phoneme (i.e., the

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

syllabic onset) is considered: thus, 215 of 217 nouns ending with C + *-on* are masculine if the consonant is neither *s* nor *ç* (e.g., *le violon*). The endings *-son* and *-çon* account for 68 nouns and are considered ambiguous (e.g., *un glaçon, une leçon*). Noteworthy, however, is that if derivational suffixes were used to operationalise noun endings, *-aison* would prove to be a reliable predictor of gender, with all 17 nouns with this ending in the corpus being feminine (e.g., *la liaison*).

The final phoneme /e/, according to Table 1, is the least reliable of all final phonemes, being more or less split between feminine and masculine genders. Its orthographic representations, however, are for the most part reliable predictors. The ending *-é*, on its own, is indeed ambiguous, but is able to predict gender when considered in conjunction with preceding phonemes: that is, the syllable *-té* predicts feminine gender (e.g., *la quantité*) whereas syllables consisting of *-é* preceded by any consonant other than *t* predict masculine gender (e.g., *un pavé*). Allowing for this exception in the analysis, of the 491 nouns ending with *-ée* (e.g., *une année*) and *-té* (e.g., *une volonté*), 95 per cent are feminine, whereas of the 493 nouns ending with either *-er* (e.g., *le danger*) or *-é* after a consonant other than *t* (e.g., *le passé*) plus 4 with *-ed* (e.g., *le pied*), 99.6 per cent are masculine. The diphthongal nucleus *-ié* is considered ambiguous, unless animate nouns are distinguished from inanimate nouns, in which case all 4 inanimate nouns are feminine (e.g., *la pitié*) and all 9 animate nouns are masculine (e.g., *un allié*).

Also identified in Table 3 are two final phonemes whose orthographic representations reliably predict either feminine or masculine gender without any ambiguous endings to account for. In the case of the final phoneme /i/, 98 per cent of the 358 nouns with feminine endings (primarily *-ie*: *une vie, une biographie*) are feminine, while 97 per cent of the 165 nouns with masculine endings (e.g., *-i*: *un cri*; *-is*: *un colis*) are masculine. Similarly, the distribution of orthographic endings with /y/ as their final phoneme shows that 100 per cent of 31 nouns ending with the feminine ending *-ue* are feminine (e.g., *une avenue*) and that 94 per cent of 64 nouns with masculine endings (e.g., *-u*: *un reçu*) are masculine. Finally, in the case of /a/, of 135 nouns with orthographic endings identified as masculine (e.g., *-as*: *un matelas; -at*: *un état*), 99 per cent are masculine; 8 of 9 nouns ending with *-oie* are feminine and so this ending is classified here as feminine even though its predictive value of 89 per cent falls just below the minimum of 90 per cent. The remaining 115 nouns have orthographic endings that are considered ambiguous (e.g., *-ia*: *une cafétéria, un pétunia*).

3.2 *Consonantal final phonemes*

Table 4 identifies orthographic endings of consonantal final phonemes that are predominantly feminine. The majority of the 598 nouns ending in /s/ are feminine and include a set of orthographic endings that reliably predict feminine gender for 93 per cent of 484 nouns (e.g., *-ance*: *la distance*). Another set of orthographic endings predict masculine gender for 99 per cent of 77 nouns (e.g., *-us*: *un cactus*), while 37 nouns have ambiguous endings that do not predict gender.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 4. *Orthographic endings with consonantal final phonemes that are predominantly feminine*

Final phoneme	Orthographic endings	Number of nouns by gender F	M	F/M	Total	Predictive value
/s/	Fem (*classe*): *-asse, -ace, -esse, -ète, -aisse, -isse/-ice, -uce, -ousse, -ance, -anse, -ence, -ince, -once, -ource/-ourse*	451	30	3	484	93 per cent
	Masc (*atlas*): *-as -ès, -is, -ils, -os, -us, -ess, -ex, -yx, -inx/-ynx*	1	76	0	77	99 per cent
	Amb (*force/divorce*): *-arce/-arse, -erce/-erse, -orce/-orse, -ipse/-ypse, -axe, -exe, -ixe, -oxe*	21	15	1	37	–
/n/	Fem (*antenne*): *-anne, -enne, -onne, -une, -ine, -aine, -eine, -erne, -urne*	243	7	6	256	94 per cent
	Masc (*pollen*): *-en*	0	9	0	9	100 per cent
	Amb (*tisane/butane*): *-ane, -ène, -one, -oine*	36	39	8	83	–
/d/	Fem (*viande*): *-ande, -ende, -onde, -ade, -ude, -aude, -arde, -orde, -ourde*	173	6	4	183	95 per cent
	Amb (*méthode/code*): *-ède, -ide, -ode*	14	14	19	47	–
/z/	Fem (*perceuse*): *-euse, -ouse, -ase, -aise, -èse, -oise, -ise, -yse, -ose, -ause, -use*	229	5	1	235	97 per cent
	Amb (*topaze/trapèze*): *-aze/-èze*	2	2	0	4	–
/ʃ/	Fem (*moustache*): *-ache, -âche, -iche, -èche, -êche, -oche, -uche/-ûche, -auche, -ouche, -arche, -erche, -orche, -anche/-enche*	92	6	4	102	90 per cent
	Masc (*flash*): *-ash*	0	3	0	3	100 per cent
/v/	Fem (*cave*): *-ave, -ève, -ive, -ouve, -erve, -alve*	51	2	4	57	90 per cent
	Amb (*preuve/fleuve*): *-êve, -uve, -auve, -euve*	6	5	0	11	–

The following patterns emerge for the final phoneme /n/:

- a set of feminine endings (e.g., *-enne, -ine*) predict feminine gender for 94 per cent of the 256 nouns with these endings (e.g., *la farine*);
- nouns ending with *-en* are masculine (e.g., *un specimen*);
- *-ane, -ène, -one,* and *-oine* predict neither gender (e.g., *une zone, un trombone*).

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 5. *Orthographic endings with consonantal final phonemes that are predominantly masculine*

Final phoneme	Orthographic endings	Number of nouns by gender F	M	F/M	Total	Predictive value
/ʒ/	Masc (*fromage*): *-age, -ège*	6	237	2	245	97 per cent
	Amb (*grange/change*): *-(V)uge, -ige, -eige, -oge, -erge, -orge, -ange, -inge, -onge*	30	26	2	58	–
/m/	Masc (*tourisme*): *-ème, -ome/-ôme, -aume, -yme, -asme, -isme, -ysme, -am, -em, -um, -ilm*	3	156	4	163	96 per cent
	Amb (*plume/légume*): *-amme, -emme, -omme, -ame, -ime, -ume, -alme, -arme, -erme, -irme, -orme*	43	40	3	86	–
/f/	Masc (*motif*): *-if, -ef, -euf*	1	50	1	52	96 per cent
	Amb (*autographe/paragraphe*): *-afe, -iffe, -ophe, -aphe*	9	6	4	19	–
/ŋ/	Masc (*camping*): *-ing*	0	7	0	7	100 per cent

Noteworthy is that morphology, if taken into account, does play a small role here in predicting masculine gender for seven nouns ending with the suffix *-phone* (e.g., *un magnétophone*) and four ending with the suffix *-gone* (e.g., *un polygone*).

Orthographic endings of 230 nouns whose final phoneme is /d/ include feminine endings (e.g., *-ande, -ade*) that predict the gender of 95 per cent of 183 nouns (e.g., *une amande*), while *-ède, -ide,* and *-ode* are considered ambiguous (e.g., *une ride, un fluide*). With the exception of only four nouns with ambiguous endings, the orthographic endings of 239 nouns whose final phoneme is /z/ are strong predictors of feminine gender, with feminine endings (e.g., *-euse, -ise*) predicting the gender of 97 per cent of 235 nouns (e.g., *une agrafeuse, une cerise*). Of 105 nouns whose final phoneme is /ʃ/, 103 have feminine endings such as *-ache* and *-oche* that predict feminine gender for 90 per cent of nouns with those endings (e.g., *une poche*), while the three nouns ending in *-ash* (e.g., *un flash*) are masculine. Finally, 68 nouns contain the final phoneme /v/: of 57 nouns with feminine endings (e.g., *-ave, -eve, -ive*), 90 per cent are feminine (e.g., *une betterave*); the remaining 11 nouns have orthographic endings considered ambiguous (e.g., *-euve*: *une preuve*, *un fleuve*). Table 5 identifies orthographic endings of consonantal final phonemes that are predominantly masculine. The most frequent of these is /ʒ/, which accounts for 303 nouns. Nouns ending in *-age* and *-ège* are both relatively numerous and predictive of masculine gender: 97 per cent of 245 nouns with these endings are masculine (e.g., *un garage*). The remaining 58 nouns ending in /z/ are ambiguous in terms of the predictive value of rhymes (e.g., *-ige*: *la voltige*, *le prestige*). In the case

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

of /m/, a set of masculine endings (e.g., *-ème, -isme, -um*) predicts masculine gender for 96 per cent of 163 nouns (e.g., *un poème, le tourisme, un album*). Another set of endings is ambiguous and does not predict the gender of the other 86 nouns ending with /m/ (e.g., *-ime: une cime, un crime*). With regard to the 71 nouns with /f/ as their final phoneme, masculine endings such as *-if* occur in 52 nouns (e.g., *le récif*), 96 per cent of which are masculine. Ambiguous endings occur in the remaining 19 nouns. Finally, of the small number of nouns whose final phoneme is /ŋ/, all 7 end with *-ing* and are masculine (all borrowings from English, such as *le camping*).

Table 6 and Table 7, both of which display orthographic endings with consonantal final phonemes that are not strongly associated with only one gender or the other, have been separated, simply to facilitate presentation, according to the number of nouns with these endings. Table 6 displays consonantal final phonemes that occur in the corpus with medium to high frequency. The final phoneme /R/ is the most frequently occurring final phoneme in the entire corpus, affecting about 15 per cent of all nouns. At first glance, /R/ as a final phoneme appears to be unreliable as a predictor of grammatical gender, ranking low in Table 1, indicating masculine gender for 63 per cent of the 1,507 nouns ending with /R/. Again, however, orthographic endings increase predictive value, as shown in Table 6. For example, 96 per cent of the 302 nouns with orthographic endings identified as feminine are feminine—the most frequent feminine endings being *-ure* as in *une blessure* and *-ière* as in *une carrière*. With respect to orthographic endings identified as masculine, 97 per cent of 572 nouns with these endings are masculine, one of the most frequent being *-eur* in the case of animate nouns. Also relatively frequent and 100 per cent reliable are the following endings: *-oir, -ir, -or*, and *-ar*. However, 633 nouns are considered to have ambiguous endings (e.g., *-aire*: *une grammaire, un vocabulaire*; *-ère*: *une fougère*, *un mystère*). Helpful to note, though, is that all words ending in *-oir* and *-air* are masculine, whereas their homophonous counterparts *-oire* and *-aire* are ambiguous endings unable to predict gender. Orthographic endings with /t/ as their final phoneme also appear in Table 6 along with their distribution across 679 nouns. Orthographic endings identified as feminine (e.g., *-ente: une tente*) predict feminine gender for 95 per cent of 393 nouns with these endings. Orthographic endings identified as masculine (e.g., *-it, -ut*) predict masculine gender for all 26 nouns (e.g., *le granit, le scorbut*). Several orthographic endings with /t/ as their final phoneme are considered ambiguous, unable to predict the gender of 260 nouns (e.g., *-ite*: *une visite, un site*; *-ote*: *une note, un vote*). The high proportion of ambiguity is due in part to the large number of animate nouns that end with *-iste* and have dual gender (e.g., *un journaliste, une journaliste*).

Analysis of 561 nouns with /l/ as their final phoneme reveals a majority of ambiguous endings: namely, 299 nouns ending either in *-le* after V (e.g., *une casserole, un symbole*) or *-le* after V(C)C if C $\neq$ *l* (e.g., *une boucle, un cercle*). More reliable predictors of grammatical gender are the orthographic endings identified as feminine (e.g., *-elle*: *une voyelle*), which predict gender for 92 per cent of 101 nouns, in addition to orthographic endings identified as masculine (e.g., *-al*: *un hôpital*), which predict the gender of 99 per cent of 161 nouns.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 6. *Orthographic endings with medium- to high-frequency consonantal final phonemes that are feminine, masculine, or ambiguous*

Final phoneme	Orthographic endings	Number of nouns by gender				Predictive value
		F	M	F/M	Total	
/R/	Fem (*lumière*): *-ière, -ure, -eure*	290	10	2	302	96 per cent
	Masc (*éclair*): *-air, -er, -erf, -ert, -ur, -eur* (if animate), *-ar, -arc, -ars, -art, -our, -ours, -ourd, -or, -ord, -ors, -ort, -aure, -ir, -oir*	12	556	4	572	97 per cent
	Amb (*sphère/critère*): *-ère, -eur* (if inanimate), *-are, -ore, -oire, -ire, -yre, -ydre, -re* after V(C)C	210	364	59	633	–
/t/	Fem (*dette*):, *-ette, -ête, -ète, -atte, -aite, -otte, -utte, -aute, -eute, -oite, -oute, -arte, -erte, -orte, -ante, -ente, -inte, -onte*	375	10	8	393	95 per cent
	Masc (*granit*): *-it, -ut, -out, -act, -ict, -est, -omte, -ompte*	0	26	0	26	100 per cent
	Amb (*date/nitrate*): *-ate, -ite, -ote, -ute, -acte, -ecte, -alte, -olte, -ulte, -aste, -este, -iste, -oste, -yste, -ouste, -epte, -ipte/-ypte*	99	47	114	260	–
/l/	Fem (*salle*): *-alle, -elle, -ille, -olle, -ulle, -ylle*	93	6	2	101	92 per cent
	Masc (*festival*): *-al, -el, -il, -ol, -ul, -eul, -all*	1	160	0	161	99 per cent
	Amb (*céréale/scandale*): *-ale, -èle/-êle, -ile, -ole, -(V)ule, -le* after V(C)C (C≠*l*)	151	125	23	299	–
/k/	Masc (*lac*): *-ac, -ec, -ic, -oc, -uc, -ak, -ock*	0	51	0	51	100 per cent
	Amb (*clinique/portique*): *-aque, -èque, -ique, -oque, -uque, -anque, -arque, -orque, -asque, -esque, -isque*	85	43	2	130	–
/j/	Fem (*famille*): *-ille, -aille, -eille, -ouille, -aye*	93	6	1	100	93 per cent
	Masc (*travail*): *-ail, -eil, -euil, -ueil*	0	43	0	43	100 per cent

The majority of 181 nouns with /k/ as a final phoneme also have ambiguous endings, all ending in *-que* (e.g., *-ique*: *une boutique*, *un graphique*). However, 100 per cent of the 51 nouns identified as having masculine orthographic endings are

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 7. *Orthographic endings with low-frequency consonantal final phonemes that are feminine, masculine, or ambiguous*

Final phoneme	Orthographic endings	Number of nouns by gender				Predictive value
		F	M	F/M	Total	
/p/	Fem (*nappe*): *-appe, -oppe, -ampe/-empe, -ompe, -arpe*	18	1	0	19	95 per cent
	Masc (*cap*): *-ap, -op*	0	5	0	5	100 per cent
	Amb (*coupe/groupe*): *-ape/-âpe, -ète/-êpe, -ipe/-ype, -ope, -upe, -oupe*	24	17	1	42	–
/g/	Fem (*bague*): *-ague, -igue, -(V)ugue, -angue, -ingue*	19	0	1	20	95 per cent
	Masc (*boomerang*): *-ang, -ong*	0	4	0	4	100 per cent
	Amb (*synagogue/catalogue*): *-ègue, -ogue, -orgue*	5	8	17	30	–
/b/	Fem (*barbe*): *-arbe, -ombe, -ourbe*	10	0	0	10	100 per cent
	Masc (*tube*): *-ube, -ub*	0	4	0	4	100 per cent
	Amb (*robe/globe*): *-abe, -obe, -erbe*	4	8	0	12	–
/ɲ/	Amb (*vigne/signe*): *-agne, -igne/-ygne, -ogne, -aigne/-eigne, -argne*	18	8	0	26	–

masculine (e.g., *-ac: un hamac*). Finally, among orthographic endings having /j/ as their final phoneme, 93 per cent of 100 nouns with feminine endings (e.g., *-aille*: *une taille*) are feminine, while 100 per cent of 43 nouns with masculine endings (e.g., *-ail*: *un détail*) are masculine.

Table 7 displays orthographic endings with low-frequency consonantal final phonemes that are not strongly associated with one gender or the other. Beginning with orthographic endings affecting 66 nouns whose final phoneme is /p/, of 19 nouns identified as having feminine endings (e.g., *-appe, -ampe*), 95 per cent are feminine (e.g., *une lampe*), and 5 nouns ending with *-ap* or *-op* (e.g., *un cap*) are masculine. The orthographic endings of the remaining 42 nouns are considered ambiguous (e.g., *-oupe*: *une loupe*, *un groupe*). Orthographic endings having /g/ as their final phoneme account for 54 nouns. Of the 20 with feminine endings (e.g., *-igue*), 95 per cent are feminine (e.g., *une digue*); 4 have masculine endings (e.g., *-ang*: *un mustang*), and 30 have ambiguous endings (i.e., *-ègue, -ogue, -orgue*). The final phoneme /b/ affects only 26 nouns in the RJ corpus. Of these, all 10 ending with *-ombe, -arbe,* or *-ourbe* are feminine (e.g., *une bombe*), the 4 ending with *-ub* or *-ube* are masculine (*un tube, un club*), while the remaining 12 have ambiguous endings (e.g., *-obe*: *une robe, un globe*). Finally, orthographic endings of the 26 nouns with /ɲ/ as a final phoneme (e.g., *-igne*) are all considered ambiguous (e.g., *une ligne, un signe*).

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 8. *Distribution of nouns in RJ corpus by ending and gender*

	Feminine gender	Masculine gender	Both genders	Total
Feminine endings	3,762	135	38	3,935
Masculine endings	61	3,998	19	4,078
Ambiguous endings	817	886	255	1,948
Total	4,630	5,019	312	9,961

3.3 Summary

To synthesize the results, Table 8 displays the distribution of all 9,961 nouns in the RJ corpus by orthographic ending (feminine, masculine, or ambiguous) and gender (feminine, masculine, or both). Read horizontally, the table reveals that:

- orthographic endings identified as feminine predict feminine gender for 96 per cent of all nouns with these endings (i.e., 3,762 of 3,935 nouns with feminine endings are feminine);
- orthographic endings identified as masculine predict masculine gender for 98 per cent of all nouns with these endings (i.e., 3,998 of 4,078 nouns with masculine endings are masculine).

Read vertically, Table 8 shows that 81 per cent of all feminine nouns in the corpus (3,762 of 4,630 nouns) are rule governed, having feminine endings that reliably predict feminine gender. Only 1 per cent of feminine nouns have masculine endings (e.g., *une dent, la peau, la fin*), while 18 per cent have ambiguous endings. Similarly, 80 per cent of all masculine nouns in the corpus (3,998 of 5,019 nouns) are rule governed, having masculine endings that reliably predict masculine gender. Only 2 per cent of masculine nouns have feminine endings (e.g., *un squelette, le silence, le braille*), while 18 per cent have ambiguous endings. Furthermore, if we exclude the 312 nouns in the corpus having both genders, we find that, of all 9,649 nouns with only one gender, 7,760 or 80 per cent have endings that reliably predict their gender.

4 CONCLUSION

This corpus-based analysis of 9,961 French nouns identified nearly 400 noun endings as either feminine, masculine, or ambiguous. The analysis revealed that 81 per cent of all feminine nouns and 80 per cent of all masculine nouns in the corpus are rule governed, having endings that systematically predict their gender.

4.1 Theoretical considerations

The most comprehensive way of documenting the effects of word-internal structural properties on gender attribution is to operationalise noun endings as orthographic representations of rhymes, consisting of either a nucleus or a nucleus

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

and a coda. When rules are derived in this way from orthographic representations of rhymes rather than from final phonemes alone, the gender attribution of an unexpectedly high proportion of nouns is rule governed and predictable. Using the 90 per cent criterion for predictive value, orthographic representations of rhymes predict grammatical gender for 80 per cent of the entire corpus, whereas final phonemes predict grammatical gender for only 18 per cent of the corpus. Final syllables were appealed to in only two cases (*-on* and *-é* after onset) in order to allow reliable patterns to emerge: the monophthongal rhyme *-on* (but not the diphthong *-ion*) generally predicts masculine gender, but is unable to predict gender reliably if preceded by *s* or *ç*; the rhyme *-é* is unable to predict gender on its own but predicts feminine gender if preceded by *t* and predicts masculine gender if preceded by any consonant other than *t*.

Final phonemes and orthographic endings interact in gender attribution, but orthography generally overrides purely phonological representations in terms of predictive strength. For example, the final vocalic phonemes /ɛ/, /Ø/, and /u/ are generally good predictors of masculine gender, with predictive values of 93 per cent, 88 per cent and 87 per cent, respectively, but all nouns ending in *-aie*, *-eue*, and *-oue* are feminine and all nouns ending in *-ai*, *-eu*, and *-ou* are masculine. Similarly, with respect to noun endings with consonantal final phonemes, the rhyme /ɛR/ is unable to predict gender attribution, yet some of its orthographic representations predict masculine gender categorically (e.g., *-air: un air; -er: un hiver; -ert: un désert*) while others are ambiguous (e.g., *-aire: une molaire, un salaire; -erre: une serre, le tonnerre*). Likewise, the rhyme /ɛl/ has no predictive value, yet *-elle* predicts feminine gender (e.g., *une ficelle*), *-el* predicts masculine gender (e.g., *un manuel*), while *-èle* is ambiguous (e.g., *la grèle, le zèle*). In still other cases, some orthographic endings obviate the need to appeal to phonology at all to explain gender attribution, as they predict gender irrespective of their final phonemes. For example, all nouns ending in *-ac, -oc, -as, -ès, -os, -us, -it,* or *-ut* are masculine, regardless of whether the final phoneme is vocalic (as in *tabac, broc, bras, succès, dos, jus, conflit, statut*) or consonantal (as in *lac, choc, atlas, palmarès, os, virus, transit, scorbut*).

Of the noun endings identified in this analysis, many, but not all, are derivational suffixes, including *-esse, -eur, -ie, -isme, -iste, -age, -ure, -ance, -ade, -ée, -ant,* and *-oir* (see Wise, 1997: 107). Therefore, derivational suffixes could arguably have been used to characterise many noun endings and to invoke morphological rules governing grammatical gender. However, orthographic representations of rhymes proved instead to be considerably more productive than derivational suffixes in their capacity to account for a greater number of nouns and to generate a comprehensive set of rules governing gender attribution. For example, productive suffixes such as *-elle and -ette* are reliable predictors of gender in nouns such as *ruelle* and *statuette*, but are even more productive as predictors of grammatical gender if not restricted to their morphological status as suffixes. That is, nouns derived from *-elle* and *-ette* are feminine (*une passerelle, une devinette*), but so are monosyllabic nouns in which these same endings constitute rhymes rather than derivational suffixes (e.g., *une pelle, une dette*). Similarly, masculine nouns derived from *-ment* were effectively subsumed in

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

the foregoing analysis under *-ent*, an orthographic representation of the nucleus /a)/ that accounts for both non-derived forms such as *talent* and derived forms such as *déménagement*, all of which conform to a single rule. Overall, the present study illustrates that final phonemes and derivational suffixes alike are limited, in comparison to orthographic representations of rhymes, in their capacity to generate a comprehensive set of rules that predict gender attribution in a corpus as large as RJ.

4.2 Pedagogical considerations

French grammars do of course address grammatical gender, but do so by focusing primarily on gender agreement, as they convey transformational rules governing differences between masculine and feminine forms of adjectives (*nouveau* ⇒ *nouvelle*) and animate nouns (*gardien* ⇒ *gardienne*), while focusing less on gender attribution.[5] Yet increasing L2 learners' knowledge of gender attribution is arguably far more efficient than drawing their attention to gender agreement rules, because the latter can be bypassed in comprehension and applied in production only once knowledge of gender attribution can be easily accessed. In other words, knowledge of gender attribution needs to precede knowledge of gender agreement rules. Yet, whereas students' attention is systematically drawn to differences between feminine and masculine adjectives such as *nouveau* and *nouvelle* to illustrate how grammatical gender affects the formation of adjectives, their attention is not typically drawn to how these same structural differences affect the gender of inanimate nouns such as *château* and *chandelle* or *rideau* and *rondelle*. Thus, as an effective starting point, existing material on transformational rules governing adjectival inflections can be extended to illustrate the similarities between adjective and noun endings according to gender. Students will benefit from having their attention drawn not only to differences between feminine and masculine adjectives (e.g., *cruel/cruelle, alpin/alpine, sain/saine, secret/secrète*), but also to how these same patterns can be observed in the gender of inanimate nouns (i.e., *le sel/la pelle, le jardin/la comptine, le pain/la laine, le jouet/la planète*).

Arguably, nearly 400 orthographic endings are too numerous to be of much use to L2 learners. The numerous endings are the result of a detailed analysis that was contingent on quantification and thus designed to account comprehensively for all nouns in the RJ corpus, excluding only proper names, numbers, some compounds, plural nouns, and nouns with unique endings. The number of endings can nonetheless be reduced for pedagogical purposes by excluding endings from the corpus that occur in, say, fewer than five nouns. This is the case in Tables 9 and 10, which display vocalic and consonantal endings that are typically feminine and typically masculine, respectively, but that include only endings that appear in at least five nouns in the RJ corpus. Exclusion of endings occurring in fewer than five

[5] Gender agreement is a variable characteristic of modifiers that is systematically derivative and depends on the context of occurrence of modifiers with nouns, whereas grammatical attribution is an inherent feature of nouns that remains invariable and independent of the context of occurrence (Carroll, 1989, 1995).

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 9. *Typically feminine endings*

Vocalic
- *-aie, -oue, -eue, -ion, -té, -ée, -ie, -ue*

Consonantal
- *-asse, -ace, -esse, -èce, -aisse, -isse/-ice, -ousse, -ance, -anse, -ence, -once*
- *-enne, -onne, -une, -ine, -aine, -eine, -erne*
- *-ande, -ende, -onde, -ade, -ude, -arde, -orde*
- *-euse, -ouse, -ase, -aise, -èse, -oise, -ise, -yse, -ose, -use*
- *-ache, -iche, -èche, -oche, -uche, -ouche, -anche*
- *-ave, -ève, -ive*
- *-ière, -ure, -eure*
- *-ette, -ête, -ète, -atte, -otte, -oute, -orte, -ante, -ente, -inte, -onte*
- *-alle, -elle, -ille, -olle*
- *-aille, -eille, -ouille*
- *-appe, -ampe, -ombe*
- *-igue*

Table 10. *Typically masculine endings*

Vocalic
- *-an, -and, -ant, -ent, -in, -int, -om, -ond, -ont, -on* (but not after *s/ç)*
- *-eau, -au, -aud, -aut, -o, -os, -ot*
- *-ai, -ais, -ait, -ès, -et*
- *-ou, -oût, -out, -oux*
- *-i, -il, -it, -is, -y*
- *-at, -as, -ois, -oit*
- *-u, -us, -ut, -eu*
- *-er, -é* after *C* (C≠t*)*

Consonantal
- *-age, -ège, -ème, -ome/-ôme, -aume, -isme*
- *-as, -is, -os, -us, -ex*
- *-it, -est*
- *-al, -el, -il, -ol, -eul, -all*
- *-if, -ef*
- *-ac, -ic, -oc, -uc*
- *-am, -um, -en*
- *-air, -er, -erf, -ert, -ar, -arc, -ars, -art, -our, -ours, -or, -ord, -ors, -ort, -ir, -oir, -eur* (if animate)
- *-ail, -eil, -euil, -ueil*
- *-ing*

nouns shortens otherwise exhaustive lists and results in more useful mnemonic aids that allow learners to home in on endings that predict gender in a productive rule-based fashion. The number of endings can be further reduced by conflating, on the one hand, homophonous vocalic rhymes (e.g., *-an, -and, -ant, -ent*) that together are reliable predictors of gender attribution and, on the other, many consonantal endings such as *-alle, -elle, -ille, -olle,* and *-ulle* that can be conveniently subsumed under '*-lle* after V' as a single ending.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Table 11. *Typically ambiguous endings*

Vocalic
• *-a, -ia, -oi/-oix, -çon/-son*
Consonantal
• *-erce/-erse, -orce/-orse*
• *-aque, -ème, -ique, -oque*
• *-ane, -ène, -one, -oine*
• *-amme, -omme, -ame, -ime, -ume, -arme, -erme, -orme*
• *-uge, -ige, -erge, -ange, -onge*
• *-erbe, -ape, -ipe, -ope, -oupe, -aphe*
• *-are, -ore, -oire, -ir, -ère, -re* after V(C)C, *-eur* (if inanimate)
• *-ède, -ide, -ode*
• *-ate, -ite, -ote, -ute, -ecte, -ulte, -este, -iste, -epte*
• *-ale, -èle/-êle, -ile, -ole, -(V)ule, -le* after V(C)C (C≠*l*),
• *-agne, -igne, -ogne, -aigne/-eigne*

Typically ambiguous endings, also excluding endings that occur in fewer than five nouns, are displayed in Table 11, not so much because learners need to increase their awareness of them, but rather to illustrate the number of ambiguous endings with a final mute *e*. Otherwise, comparison of the feminine and masculine endings in Tables 9 and 10 would lead one to validate the oft-cited rule of thumb whereby noun endings with a final mute *e* are feminine. Most feminine endings identified in the corpus do indeed end with mute *e*, whereas considerably fewer masculine endings contain a final mute *e* (*-isme* and *-age* being the most noteworthy). However, when the entire corpus is considered, including ambiguous endings, the rule of thumb proves to be unreliable. Overall, noun endings with a final mute *e* indiscriminately include either feminine or ambiguous endings, with only a small proportion accounting for masculine endings. A rule of thumb that proves more reliable, therefore, is that noun endings without a final mute *e* are strong predictors of masculine gender (notable exceptions being *-té* and *-ion*).

For pedagogical purposes, nouns endings can be selected on the basis of their frequency and predictive value and then targeted through form-focused instruction. For example, in a recent classroom study designed to draw the attention of 10–11-year-old immersion students to noun endings as reliable predictors of gender attribution, Lyster (2004b) targeted a set of feminine endings (*-ie, -ion, -té, -ance, -ence, -ette, -aite, -otte, -aine, -eine, -ine, -elle, -ure,* /ʃ/, /z/) and a set of masculine endings (/ã/, /ɛ̃/, /o/, *-age*), which had been selected on the combined basis of their high predictive value and their frequency in the children's regular curriculum materials. The preponderance of feminine endings reflected the finding that learners of French L2 tend to over-assign the masculine unmarked form (Harley, 1998; Spilka, 1976). Lyster (2004a) reported that, prior to the instructional intervention, students were about three times more likely to accurately assign masculine gender than to accurately assign feminine gender, suggesting that learners needed their attention drawn especially to marked feminine forms as a means of inciting them to notice grammatical gender distinctions.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

How can students' awareness of predictive nouns endings be effectively increased? In Lyster's (2004b) study, noticing activities required students to read curriculum-related texts in which endings of target nouns were made to appear salient by means of typographical enhancement; awareness activities then required students to group nouns with similar endings together and determine which gender their endings predict; and controlled practice activities required students to accurately assign gender-specific articles to target nouns in curriculum-related materials and language games involving crossword puzzles, rhymes and riddles. Results revealed that the treatment groups receiving form-focused instruction demonstrated significant long-term improvement on two oral production tasks and a binary-choice test at the time of delayed post-testing, and showed short-term improvement on a text-completion task at the time of immediate post-testing.

Harley (1998) also conducted a study in immersion classrooms, with 7–8-year-old students in second grade, using form-focused activities designed to draw attention to noun endings that reliably predict grammatical gender (e.g., /o/, /ã/, /ɛ̃/, /ɛt/, /ɛz/, /as/, /Øz/). In her study, noticing activities required students to attend to the co-occurrence of nouns with gender-specific articles on identification labels displayed around the classroom; awareness activities required students to each create their own gender-specific dictionaries, one for masculine and other for feminine entries; and controlled practice activities required students to recall and associate nouns with similar endings in games such as 'Concentration' and to associate gender-specific articles with target nouns in various games such as 'Bingo' (to win, students had to name all objects in their winning row, accurately using gender-specific articles) and 'La valise de ma Tante' (to play, students had to name various objects with similar endings and correctly assign grammatical gender). The study demonstrated that students made significant long-term progress as demonstrated by three of the four measures (two aural-discrimination tasks and an oral picture-description task).[6] Classroom intervention studies have also been conducted with older learners (Tucker *et al.*,1977; Warden, 1998), yielding similarly successful results.

One often hears that rules of gender attribution, if they exist at all, are too unreliable to teach and to learn, because there are far too many exceptions. Yet typical French grammar rules abound with exceptions, without usually having the effect of usurping their place in reference books and grammars; see, for example, *Le bon usage* (Grevisse, 1986) or the *Complete guide to conjugating 12,000 French verbs* (1991) published in the Bescherelle series by Hurtubise. In the case of gender attribution rules, however, the present study revealed that relatively few exceptions emerge if rules are derived from noun endings that predict gender attribution at least 90 per cent of the time. At this rate, there are certainly fewer exceptions to gender attribution rules than one can find in other grammatical subsystems. Consider, for

[6] The only measure that did not reveal significant improvement was an oral task requiring students to identify the gender of low-frequency unfamiliar nouns, such as *tondeuse*, *barbotte*, and *sabot*.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

example, the following rules (and number of exceptions) for pluralising a set of irregular singular nouns in French:

- Nouns ending with *-ail* (e.g., *un travail, un détail*) usually add *s* (*les détails*), but 7 of 20 nouns (35 per cent) change *-ail* to *-aux* as in *les travaux*;
- Nouns ending with *-ou* (e.g., *un clou, un genou*) usually add *s* (*les clous*), but 7 of 33 nouns (21 per cent) add *x* as in *les genoux*;
- Nouns ending in *-al* (e.g., *un journal, un festival*) usually change *-al* to *-aux* (*les journaux*), but 9 of 62 nouns (15 per cent) add *s* as in *les festivals*.

In these cases, there are no formal rules that reliably predict one ending or the other. The learner needs to associate the correct plural form with each singular form on an item-by-item basis. In contrast, the gender attribution rules presented here are much more systematic and reliable. For example, the gender attribution rules pertaining to the same set of endings (*-ail, -ou, -al*) apply categorically to all singular nouns with these endings. That is, of the 115 nouns in the RJ corpus ending with *-ail*, *-ou*, or *-al*, 100 per cent are masculine. Even in the many cases where gender attribution rules do not apply so categorically, the predictive value of rule-governed noun endings was set at nothing less than 90 per cent. Making students aware of such reliable patterns first, and leaving exceptions for later, will enable them to internalise rules more effectively than immediately drawing their attention to exceptions.

Gender attribution rules based on noun endings, given their reliability and systematicity, are worthy of more attention in French reference books and French L2 classrooms. The foregoing corpus-based study confirmed that predictive rules for gender attribution do exist and apply to as many as 80 per cent of the nearly 10,000 nouns included in the analysis. More importantly, classroom studies have demonstrated that gender attribution rules are both teachable and learnable. Regardless of age, L2 learners can benefit from form-focused instructional activities that promote awareness of gender attribution rules and that provide opportunities for practice in associating grammatical gender with orthographic representations of constituent rhymes of literally thousands of nouns—both animate and inanimate alike.

Author's address:
Roy Lyster
McGill University
3700 McTavish Street
Montreal, QC Canada H3A IY2
e-mail: roy.lyster@mcgill.ca

REFERENCES

Batchelor, R. and Offord, M. (1993). *Using French: A Guide to Contemporary Usage*. Cambridge: Cambridge University Press.

Bérard, E. and Lavenne, C. (1991). *Grammaire utile du français*, Paris: Hatier.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Bosquart, M. (1998). *Nouvelle grammaire*. Montreal, QC: Guérin.
Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, 39: 535–594.
Carroll, S. (1995). The hidden dangers of computer modelling: Remarks on Sokolik and Smith's connectionist learning model of French gender. *Second Language Research*, 11: 193–205.
Complete Guide to Conjugating 12,000 French Verbs,. (1991). Montreal, QC: Hurtubise HMH.
Desrochers, A., Paivio, A. and Desrochers, S. (1989). L'effet de la fréquence d'usage des noms inanimés et de la valeur prédictive de leur terminaison sur l'identification de genre grammatical. *Revue Canadienne de Psychologie*, 43: 62–73.
Grevisse, M. (1986). *Le bon usage*. Paris-Gembloux: Duculot.
Harley, B. (1979). French gender 'rules' in the speech of English-dominant, French-dominant, and monolingual French-speaking children. *Working Papers in Bilingualism*, 19: 129–156.
Harley, B. (1998). The role of form-focused tasks in promoting child L2 acquisition. In: C. Doughty and J. Williams (eds), *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press, pp. 156–174.
Hawkins, R. and Towell, R. (2001). *French Grammar and Usage* (2nd ed.). New York: McGraw-Hill.
Holmes, V. and Dejean de la Bâtie, B. (1999). Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics*, 20: 479–506.
Huot, C. (1991). *Dictionnaire des genres*. Montréal: Éditions du Méridien.
Jacob, R. and Laurin, J. (1994). *Ma grammaire*. Montréal, QC: Les Éditions françaises.
Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language*. Cambridge: Cambridge University Press.
Lyster, R. (2004a, May). *How to predict the grammatical gender of 8000 French nouns*. Paper presented at the meeting of the American Association for Applied Linguistics, Portland, OR.
Lyster, R. (2004b). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26: 399–432.
Möhring, A. (2001). The acquisition of French by German pre-school children: An empirical investigation of gender assignment and gender agreement. *EUROSLA Yearbook*, 1: 171–193.
Müller, N. (1990). Developing two gender assignment systems simultaneously. In: J. Meisel (ed.), *Two First Languages: Early Grammatical Development in Bilingual Children*. Providence, RI: Foris Publications, pp. 194–234.
Ollivier, J. (1993). *Grammaire française, 2e édition*. Laval, QC: Études Vivantes.
Quemada, B. (1971). *An Inverse Dictionary of All Nouns Listed in the Petit Larousse* [microfilm]. Rochester, NY: Eastman Kodak.
Rigault, A. (1971). *La grammaire du français parlé*. Paris: Hachette.
Le Robert Junior Illustré,. *Édition nord-américaine* (1994). Montreal, QC: Dicrorobert.
Le Robert Junior Illustré,. *Édition CD-ROM* (1999). Paris: Bureau Van Dijk.
Séguin, H. (1969). *Les marques du genre dans le lexique de français écrit contemporain: compilation des cas et essai de classement*. Unpublished Masters Thesis. Montreal, QC: Université de Montréal.
Spilka, I. (1976). Assessment of second-language performance in immersion programs. *The Canadian Modern Language Review*, 32: 543–557.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press

Stevens, F. (1984). *Strategies for Second Language Acquisition*. Montreal, QC: Eden.
Surridge, M. (1985). Genre grammatical des composés en français. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 30: 247–271.
Surridge, M. (1986). Genre grammatical et dérivation lexicale en français. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 31: 267–283.
Surridge, M. (1995). *Le ou La? The Gender of French Nouns*. Clevedon, UK: Multilingual Matters.
Tucker, R. (1967). *The French speaker's skill with grammatical gender: An example of rule-governed behaviour*. Unpublished dissertation. Montreal, QC: McGill University.
Tucker, R., Lambert, W. and Rigault, A. (1977). *The French Speaker's Skill with Grammatical Gender: An Example of Rule-governed Behaviour*. Paris: Mouton.
Tucker, R., Rigault, A., Lambert, W. and Segalowitz, N. (1968). A psychological investigation of French speakers' skill with grammatical gender. *Journal of Verbal Learning and Verbal Behavior*, 7: 312–316.
Valdman, A. (1976). *Introduction to French Phonology and Morphology*. Rowley, MA: Newbury House.
Warden, M. (1998). *The effect of form-focused instruction on control over grammatical gender by French immersion students in Grade 11*. Unpublished dissertation. Toronto, ON: OISE/UT.
Wise, H. (1997). *The Vocabulary of Modern French: Origins, Structure and Function*. London and New York: Routledge.

https://doi.org/10.1017/S0959269506002304 Published online by Cambridge University Press