Glide Formation and its absence of Compensatory Lengthening in French *

Valentin Richard

ENS Paris-Saclay, Université Paris-Saclay, Cachan, France

October 2019

1   Introduction

Moraic theory [Hay89] appeared in the 1980s and greatly improved the comprehension of phoneme weights in syllables. This system provided a lot of right predictions about changes taking place inside and between syllables. All phonemes of the nucleus are assigned one mora to, and the same for phonemes of the coda in type 1 languages. Diphthongs and long vowels carry two moras or, depending on the language, may share a mora. But one major property is that the moraic structure stays intact even if surface modifications happen. This leads to the famous explained phenomena called Compensatory Lengthening (CL). If a mora-carrying phoneme is suppressed, then the mora is redistributed on another contiguous phoneme. For example in Latin, if a coda /s/ disappears, the previous vowel gets longer, as in /kas.nus/→/kaː.nus/. If the first vowel of a two-mora-carrying diphthong is transformed into a glide, then the next vowel must also get longer, as it is the case in Old English /ni.u/→/njuː/. Figure 1 displays a visualisation of the latter analysis.
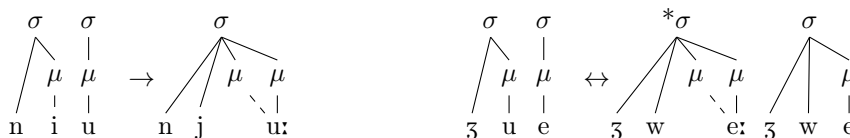

Figure 1: Visualisation of Compensatory Lengthening of /njuː/ and its absence in /ʒwe/

French also have Glide Formation with the three high vowels /i y u/ respectively identified with the three glides /j ɥ w/. For example, most French native speakers (like me) acknowledge two different pronunciations of the verb *jouer* (to play): the diaeresis /ʒu.e/ or the synaeresis /ʒwe/. Since the beginning of classical poetry it has been an endless debate to know when to count one or two syllables of such patterns [Hyp15]. The usage varies a lot between authors, but some words are decomposed differently by authors just to match the expected number of syllables of a verse, like in (1) with *ancien(ne)(s)*.

(1)   a.   (Charles Cros, Le Coffret de santal, Sonnet)
    Et, sur le char- me pris aux splen- deurs an- ci- ennes (/ɑ̃.si.jɛn/)
    1   2   3 4   5   6   7   8   9   10 11 12   .

    b.   (Charles Cros, Le Collier de griffes, En Cour d'assise)
    Les an- ciens ont dit : il fal- lait se taire (/ɑ̃.sjɛ̃/)
    1   2   3   4   5   : 6 7   8   9 10   .

Nicolas Boileau imposed the rule stating that if the two letters of the French word are present in the Latin or Greek original word, then the diaeresis must be done, otherwise the synaeresis is mandatory. This would predict two syllables for *viol* (rape) from Latin *violare* and one for *miel* (honey) from Latin *mele*. But in current Standard French (SF) both words are always pronounced in only one syllable. Moreover, the nuclear vowel is not phonemically long. Neither it is in the monosyllabic variant of *jouer*. This seems to contradict the predictions of moraic theory. This essay tries to build a model of French syllables that makes it compatible again with moraic theory. Section 2 provides data to see the distinctions to draw while speaking of Glide Formation in French.

Section 3 brings a first explanation of such an absence of Compensatory Lengthening through the prism of the way French prosody evolved. Section 4 gives the main model and section 5 implements Optimality Theory to predict the alternation with non-lengthened variants.

2   Some data

There is a one-to-on correspondence between French high vowels /i y u/ and French semi-vowels /j ɥ w/ [Wal01, DurLyc99], respectively distinguishable from one another: *miette* (crumb) /mjɛt/ - *muette* (mute.F) /mɥɛt/ - *mouette* (gull) /mwɛt/. Glide Formation (GL) can happen inside a morpheme or in between two ones (see (2)).

(2)   miel (honey) /mjɛl/   scier (to saw) /si+e/→/sje/
      lui (him) /lɥi/   ruée (rush) /ru+e/→/rɥe/
      moi (me) /mwa/   loué (rented) /lu+e/→/lwe/

However, GL is blocked by two preceding consonants in the onset of the same syllable. According to French phonotactics, it could only be an obstruent and a liquid (/l ʁ/) abbreviated *OL*. In this situation, a diaeresis in mandatory, usually creating a hiatus (3). Nevertheless, some particular patterns resist this blocking and remain tautosyllabic. This is the case of the *oi*, *oin*, *ui*, *uin* spellings respectively pronounced /wa wɛ̃ ɥi ɥɛ̃/. The *uin* case in not mentioned by the consulted authors, maybe because there are no French word containing *OLuin* not followed by a vowel. This is further investigated in appendix A.

(3)   plier (to fold) /pli.je/
      cruel (cruel) /kry.ɛl/   pluie (rain) /plɥi/
      clouer (to nail) /klu.e/   croix (cross) /kʁwa/   groin (snout) /gʁwɛ̃/

Remark that an epenthesis of a yod occurs to avoid hiatus anyway in disyllabic *iV* sequences.

The synaeresis - diaeresis alternation does not seem to be phonemic and the only minimal pairs found involve one morpheme with one of the previous mentioned particular digrams on one hand, and a polymorphemic word on the other hand (see (4)). Post-vocalic glides, on the contrary, show a great variety of minimal pairs, but only with /j/.

(4)   roi (king) /ʁwa/   roua (pummelled.3SG) /ʁu+a/→/ʁu.a/
      trois (three) /tʁwa/   troua (tore.3SG a hole in) /tʁu+a/→/tʁu.a/
      ail (garlic) /aj/   haï (hated) /a+i/→/a.i/
      abeille (bee) /a.bɛj/   abbaye (abbey) /a.be.i/
      pied (foot) /pje/   piller (to plunder) /pij+e/→/pi.je/

One other exception to GL is prefixing. Hiatus is forced when the prefix ends in a high vowel and the stem begins with a vowel, as in *dioxygène* (dioxygen) /di+o.ksi.ʒɛn/→/di.o.ksi.ʒɛn/ or *anti-âge* (antiaging) /ɑ̃.ti+aʒ/→/ɑ̃.ti.aʒ/. Moreover, a /j/ is less commonly inserted. As well, no GF is performed between stems (*tissu-éponge* (terrycloth) /ti.sy+e.põʒ/→/ti.sy.e.põʒ/) or words (*j'envie Alain* (I envy Alain) /ʒ+ɑ̃.vi+a.lɛ̃/→ /ʒɑ̃.vi.a.lɛ̃/) [Han95]. Fast speech on common expressions can sometimes trigger GF, but the only real exception to this is with the verb clitic *y*. For instance *il y a* (EXPL there.CL has: there is/are) can be said /i.lja/ and even further reduced to /ja/ with the corresponding spelling *y'a*.

In Standard French (SF), some morphemes can be pronounced either by synaeresis or by diaeresis, with some preference depending on the context (*hier* (yesterday) /i.jɛʁ/ or /jɛʁ/, but *avant-hier* (the day before yesterday) only /a.vɑ̃.tjɛʁ/). Some alternations are only heard in poetry, like (1). In casual speech, this alternation seems to be more common at the morpheme boundary. An intern sonority scale /j/</w/</ɥ/ makes the diaeresis more frequent in *nuée* (swarm) than in *noué* (knotted) than in *nié* (denied).

Conservative French spoken in the Midi (south of France) obeys different rules of GF [DurLyc99], so that *lier* (to bound) and *cendrier* (ashtray) contain both two syllables instead of one and three respectively in SF. Among Romance languages, French is the one with the most GF. Romanian, and Spanish in a lesser extend, preserve a diphthong - hiatus contrast, whereas Portuguese has only hatius [ChiHua07].

3   Historical changes and current French prosody

Some of the French diphthongs can be traced back to Old French descending from Vulgar Latin [ChiHua07]. The Vulgar Latin open /ɛ/ inherited from Classical Latin long /eː/ broke into a diphthong in several daughter languages when it was stressed. Compare French *p*[jɛ]*rre* with Spanish *p*[je]*dra* or Romanian *p*[ja]*tră*, all from Late Latin *p*[ɛ]*tra*. This is the same for *miel* and *pied*. Similarly, Classical Latin long *ē* transformed into Vulgar Latin closed /e/, transformed into the current *oi* sequence, like in *tēla*→[tejlə]→Old French [tojlə]→*toile* (cloth)

[twal]. In *doigt* (finger) from Vulgar Latin *dĭtus*, the same processus must have occurred as *ĭ* merged with /e/. The old orthography *deit* corroborates it. The lost of short - long vowel opposition tended towards a simple moraic structure for these diphthongs, as shown in 2. This predicts impossible diaeresis for this words. On the contrary, most *ui* sequences come from separate phonemes, like *huit* (eight) /ɥit/ from Latin *ŏcto*, where the /k/ palatalised in /j/, or *fuir* from *fugire*. Boileau's rule would then require a diaeresis, what is indeed found in classical literature, but also before (5).

(5)   (Pontus de Tyard, Élégie d'une dame énamourée d'une autre Dame)

     Que j'ay fu- y en vain tous les au- tres A- mours !
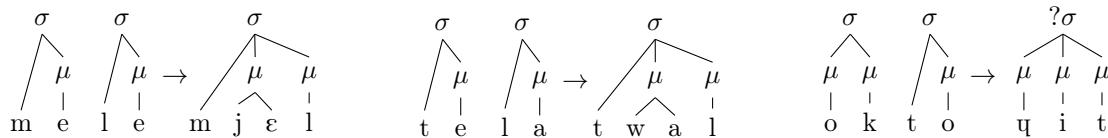   1   2   3   4 5 6   7   8 9   10   11 12   !

Figure 2: Vowel breaking (/mjɛl/ and /twal/) and GF (/ɥit/) creating diphthongs between Vulgar Latin and modern Standard French

The theory of [ChiHua07] claims that existing diphthongs in French attracted *GV* sequences to be also diphthongised. As they emerged from monomoraic nuclei under stress, the GF could have copied this structure, then avoiding Compensatory Lengthening. But introducing the glided vowel in the nucleus does not help explicating the phenomena in (3) with *OLGV* sequences (see next section). So we have to summon a more general property of French

The loss of distinction between short and long vowels lead to a unique phonemic length in French. Phonetic lengthening is only observed in stressed closed syllables (i.e. at the end of phonological phrases) if the nucleus vowel is /ø o ɑ ɛ̃ œ̃ ɔ̃ ɑ̃/ or if it is followed by /v z ʒ ʁ vʁ/ [Wal01]. Elsewhere, all syllables are of the same length. This means that it is not allowed to put two moras in the nucleus. Therefore, heavy long vowels are just forbidden.

Nevertheless, French spoken outside France conserves some long vowels, especially the /ɛ ɛː/ distinction. It is often correlated with circumflex accents, like in *fête* (party) /fɛːt/ or *maître*[1] (master) /mɛːtʁ/≠/mɛtʁ/ *mettre* (to put). The etymological spelling shed a light on this phenomenon. The circumflex accent is there to recall the former presence of an *s*, which is still there in cognate words like *festif* (festive). Coming from Old French *feste*, a Compensatory Lengthening did happen. However, this phoneme is now lexicalised and not productive. So it cannot trigger CL with GL anymore.

4   French Syllable model

The syllable theory of Klein [Kle91] argues for only two main components of the syllable: the onset and the rhyme. Each semi-cycle follows a valley-peak-valley expansion curve, i.e. the sorority is first ascending and then descending. Glides occupy the ambivalent role just at the interface of the onset and the rhyme. Therefore, they can either be treated as members of the first or of the latter. This seems to be visible in French with the opposition between consonantic glides, which prohibit liaison, and vocalic glides allowing it: (6). Namely, all *oi oin ui* are vocalic. These sequences are analysed as strong coherence diphthongs by Durand and Lyche [DurLyc99]. It means that they share a single mora. Especially, they are both in the rhyme, contrary to any other GF where the glide is assigned to the onset. Together with the impossibility to have a *OLG* sequence in the onset, it predicts exactly (3).

(6)   l'oie (the goose) /lwa/   le western (the western) /lə.wɛs.tɛʁn/
     l'oindre (to anoint him/her) /lwɛ̃dʁ/   le ouistiti (the marmoset) /lə.wis.ti.ti/
     l'huile (the oil) /lɥil/   le huitième (the eighth) /lə.ɥi.tjɛm/
     l'ouest (the west) /lwɛst/   la hiérarchie (the hierarchy) /la.je.ʁaʁ.ʃi/
     les yeux[2] (the eyes) /le+jø/→/le.zjø/   le yaourt (the yogurt) /lə.ja.uʁ(t)/

But some problem remains with vocalic glides in the onset. Klein suggests that they are kind of made invisible for French sandhi phenomena. On the contrary, I suggest that vocalic and consonantic glides occupy different

---

[1]The circumflex accent on *maître* has been deleted in the orthography reform of 1990 but is still mainly used

[2]Note that in the case of *les yeux*, the vocality of the glide can be explained only lexically, as the slang word *zyeuter* (to scrutinise) suggests.

positions in the onset, the medial or the initial place respectively. The initial contains at most a sibilant S and a consonant C. The medial H can only contain a liquid or an semi-vowel. Then the liaison is allows exactly when there is no initial. As an OL sequence already occupies C and H, a diaeresis must be done instead of a rising diphthong. Figure 3 summarises this model.
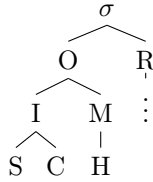

Figure 3: Partial French syllable model by distinguishing the medial M from the initial I in the onset

One other remark is about the *ui* sequence. If it acquired two moras in early Vulgar Latin, with the lost of phonemic length, one mora must have been deleted. There are two possibilities to do so: either by displacing the high vowel in the onset or by reattaching it to the mora of the following vowel. This explains the two main diphthong sorts of French: the second one being characterised as strong coherence was the way /ɥi/ has evolved, contrary to all other ones. Concerning /jɛ/ as in *miel*, we could argue that, because more other diphthongs with a yod did not come from a single vowel, the yod was detached from the vocalic mora to land in the onset. This predicts no modern distinction between strong coherence /jɛ/ and weak coherence one.

Hannahs [Han95] explains the asymmetry between GF between stem + suffix and prefix + stem or stem + stem by means of the phonological word. French phonological word (PhWd) only includes the stem and its suffixes, the rest being part of other phonological words. Yet GF is only allowed inside a PhWd. Going a bit further, the stem-suffix boundaries could be analysed as not totally permeable. This would explain why GF is not always mandatory or show variations on morphemic boundaries. Even words today considered monomorphemic, like *nuage* (cloud) from *nue+age*, show this tendency to allow or here prefer diaeresis /nu.aʒ/.

Post-vocalic glides should be treated differently in light of the data (4). The most relevant is to postulate an underlying /j/ inherited from the delateralisation of /ʎ/ but still spelled *(i)l(l)*. This yod could have been attracted from the preexisting glides and then merged. Therefore, we could contemplate a future fully distinct yod even prevocalically.

5 Optimal Theory model

In this section we recall the Optimality Theory analysis of Standard French GF by Durand and Lyche [DurLyc99] to see how it can be completed to account for the phenomena discussed above. We will only focus on glides at the morphemic boundary, like *jouer* /ʒu+e/. As CV is the preferred syllable pattern in French, we need the constraint On. violated at each syllable without onset. As well, Tri is violated by ternary syllables (having three elements in the onset or more). French prosody forbids us to use branching moras, except for so-called strong coherence diphthongs, which is summarised in *Bra-μ. This is even more restrictive between morphemes. Therefore, *Bra-AM-μ being violated at each mora branching across two morphemes has to dominate most other constraints. The only fidelity constraint required here is MaxRF stipulating that every features of the last segment on an input stem must have a corresponding element in the output. The ordering suggested by [DurLyc99] and some examples are displayed in Table 1. A dash is used to make explicit the onset-rhyme split.

/ʒu+e/	*Bra-AM-μ	Tri	*Bra-μ	On.	MaxRF
a. ʒ-(we)$_\mu$	*!				*
b. ʒu.e				*!	
☞ c. ʒwe					*
/tʀu+e/	*Bra-AM-μ	Tri	*Bra-μ	On.	MaxRF
a. tʀ-(we)$_\mu$	*!				*
☞ b. tʀu.e				*	
c. tʀwe		!*			*

Table 1: Ordering of constraints for Standard French GF with two examples

Light long vowels are avoided thanks to Bra-μ. To avoid heavy long vowels and diphthongs, we have to add another most dominating constraint: *Nuc-$\mu\mu$ prohibiting two moras in the nucleus. This forbids any Compensatory Lengthening. Then it suffices to remark that the synaeresis-diaeresis alternation is just due to

the reordering of On. and MaxRF. So, by letting the two constraints unordered we predict both surface forms. Results are given in Table 2

/ʒu+e/	*Nuc-$\mu\mu$	*Bra-AM-μ	Tri	*Bra-μ	On.,MaxRF
a. ʒ-(we)$_\mu$		*!			*
☞ b. ʒu$_\mu$.e$_\mu$					*
☞ c. ʒwe$_\mu$					*
d. ʒ-w$_\mu$e$_\mu$	*!	*			*
e. ʒw-eː$_{\mu\mu}$	*!				*
f. ʒw-eː$_\mu$				*!	*

Table 2: Minimal modification of the model to account for the studied phenomenon

In the case of *lier*, the /li.je/ surface form does not violate any constraint because an onset is reestablished on the second syllable (*Dep-IO* being weaker). Therefore it wins. But as /lje/ is still frequently used, a weighted model giving a slightly lower weigh to MaxRF should predict this better. However, it should still be understood why inserting a corresponding glide on the following onset is not common for /y u/. Maybe this is related to the strong phonemic power of /j/ compared to /ɥ w/.

6   Conclusion

Absence of Compensatory Lengthening is a feature of French synaeresis-diaeresis alternation that is explained by its history and prosody. On the one hand, the emergence of monomoraic diphthongs attracted Glide Formation. On the other hand, the loss of phonemic length and stress forbids any heavy diphthongs and long vowels. An Optimal Theory model allows us to see how these constraints arrange and correctly predicts the data.

Also in the domain of poetry debates, the count of the final *e* was claimed to be related to some same OT constraints (namely MaxSF) by [DurLyc99]. If this phenomena were to be indeed correlated, this would give a strong point in favour of OT to account for global properties of natural languages.

References

[Hay89] Hayes, B. Compensatory Lengthening in Moraic Phonology. Linguistic Inquiry, 20(2) (1989), 253-306. Retrieved from http://www.jstor.org/stable/4178626

[DurLyc99] Durand, Jacques & Lyche, Chantal. Regard sur les glissantes en français : français standard, français du Midi. Cahiers de Grammaire 24 (1999), ≪ Phonologie : théorie et variation ≫, pp. 39-65

[Han95] Hannahs, S.J.. The phonological word in French. Linguistics. Linguistics 33(6) (1995), pp. 1125-1144, DOI: 10.1515/ling.1995.33.6.1125

[ChiHua07] Chitoran, Ioana & Hualde, José. From hiatus to diphthong: The evolution of vowel sequences in Romance. Phonology 24(01) (2007), pp. 37 - 75, DOI: 10.1017/S095267570700111X

[Wal01] Walker, Douglas. French Sound Structure. University Of Calgary Press (2001), DOI: 10.2307/j.ctv6gqrmv.

[Hyp15] Hyperion Darius. Essai sur la diérèse et la synérèse. Les Amertumes (2015), http://darius.hyperion.over-blog.com/2015/07/essai-sur-la-dierese-et-la-synerese.html

[Kle91] Klein, M. Vers une approche substantielle et dynamique de la constituance syllabique. Le cas des semi-voyelles et des voyelles hautes dans les usages parisiens. Thèse de doctorat (1991), Saint Denis, Université de Paris VIII.

A The case of *uin*

The /ɥɛ̃/ sequence, as realised in *uin* followed by a consonant or the end of the word, is rare in French (*juin* (june) /ʒɥɛ̃/, suinter (to ooze) /sɥɛ̃.te/). But as *oin* follows the behaviour of *oi*, it is relevant to ask whether *uin* follows the behaviour of its denasalised equivalent *ui*. The English Wikipedia claims that /wɛ̃/ also counts as a strong coherence diphthong. However, no French word contains the *OLuin* sequence followed by a consonant or the end of the word.

Word	truie	ploué	bluin	troin	crier	psion	groin	druin	flier	cloué	cluie
Expected #syllable	1	2	1	1	2	1	1	1	2	2	1

Table 3: List of tested word in that order

My idea to test this hypothesis is to see how native speakers split this sequence in invented words. I conducted a little survey on 9 Standard French native speakers. They were asked to tell how many syllables were contained in the words in Table 3, all of the form *OLGV* (except *psion*). The words belong to three categories: A) words with weak coherence diphthongs (*oué* and *ier*), B) words with strong coherence diphthongs (*ui* and *oin*) and C) words with *uin*. As control, as many existing words (tag L: lexical) ad invented ones (tag I: invented) were put in A and B. A's lexical units are expected to receive an average #syllable rating of 2, and of 1 for B's. Results are provided in Figure 4.
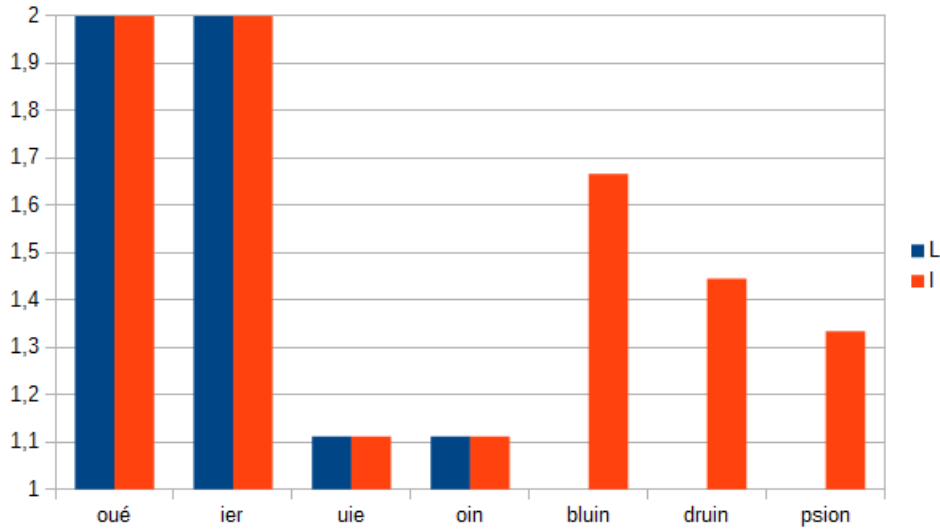

Figure 4: Average answered #syllable for each type of sequence. L: existing word, I: invented word.

The results show no clear demarcation of /ɥɛ̃/ between one or two syllables in a GF blocking context. With so few participants, differences between rating *bluin* and *druin* could be only due to a priming effect by the preceding word in the list. It seems to be the most reasonable at that point. Therefore, we cannot conclude whether *uin* is a strong coherence diphthong of French. Native speakers have maybe no judgement on that question, hence values around 1,5.

More interestingly, I added the *psion* sequence. This pattern is forbidden in a single syllable, as says [Wal01]. Every such sequence strictly inside a word would displace the /p/ in the coda of the previous syllable by enchainement. But the initial *ps* do happen is a wide range of Greek loan morphemes, like *psych+*. So we can wonder whether that sequence entered French language, and in this case, whether it behaves like *OL*. The chosen invented word (because no such initial *psGV* exists) has a weak coherence diphthong. So it should be spelled in two syllables if and only if *ps* is considered as *OL*.

The results are not marked enough and another broader experiment must be carried out. However, confirming the idea that *s* does not behave like a liquid, some predilection to monosyllabise *psion* argues for an interesting behaviour. Some further work on the structure of French syllable could take this new phenonema into account, due to the fact that is not compatible with the above mentioned models.