

Yan Greub

How to avoid the risk of transforming a great printed dictionary into a poor digitized one? Some partial answers, for want of more complete ones.

1. The FEW is a huge dictionary when we consider the sheer mass of data (25 volumes, 16000 pages) and its exhaustive aims. It has indeed the purpose of registering and etymologizing the whole lexicon, not only of French, but also of earlier stages of the language and of Occitan; of every Gallo-romance dialect; of every technical or professional genre; of every language register, including slang. Summing up, the FEW aims to include and describe every single lexical unit which exists or has existed in the territory of ancient Gaul. The term “*exist*” here is used in a very broad sense: you can find in the FEW forms that are hapaxes (i.e. one-off forms), which means that the dictionary is not solely limited to what belongs to the *langue* (in the Saussurean sense), but also includes some part of *parole* (i.e. what people as individuals actually do or did with the language, how they used it). The sheer size of this undertaking means two things:

Firstly, there is a huge amount of data; secondly, the presentation and organization of the data is exceedingly complex. Concerning the first point, there are roughly 5 million lexical units, if we follow existing estimations. Some of this data is quite easily accessed via other sources, but this is not always the case. The FEW is one of the few dictionaries where one can find historical documentation on French locutions, for example (even if the situation has improved more recently). The sheer amount of the material is, though, a fast growing problem: we are currently working on the new redaction (Neubearbeitung) of the alphabet’s beginning, and more precisely on the letter B, and the proportion of new inclusions is increasing continually: for the letter A, we rewrote the first edition of 191 pages to a new total of 1380 pages, i.e. roughly seven times the original. However, for the letter B, we predict that there will be more like 20 or 30 times the original matter.

Concerning the problem of the mass of data, we must add the fact that each item has to be etymologized according to the principles of the *etymologie-histoire*, that is not only by giving the “origin” of words, but by writing the complete history of their developments, of the meaning’s changes and of the modifications of the relations inside the lexical families. So it is not too hard to understand that, if the FEW is a dictionary of 16000 pages, this is because of its extreme concentration and the tightness of its presentation. So we come to the second issue concerning the size of the FEW: the organization of the material’s presentation. The second part of the dictionary’s full name reminds us that it aims to be a *Darstellung des galloromanischen Sprachschatzes*. It is the representation itself which draws the etymology, or the etymological position of the lexical units within the context in which they appear. This representation is designed around the organization of the articles, as I will attempt to show in what follows.

An FEW article puts together, under an entry given by an etymon, its entire descent in the Gallo-romance varieties, that is a whole lexical family. It can have from 1 lexical unit to 60 pages.

Let us see an example: the article *mundus* (appendix). The main points you can observe are the absolute segregation between the commentary and the material; the distinction (I. vs. II.) between the lexicon borrowed and naturally evolved, and the smaller divisions designing linguistic units, and more precisely historical units.

By studying a photograph of Wartburg making an article (see appendix) we can see how the etymologisation and the intellectual activity of understanding the problems exposed in an FEW article, are inherent in the disposition of the material itself. Wartburgian etymology relies on the complete exposition of the lexical material. In this sense, it is contrary to the etymology of Malkiel or Corominas, who present the pertinent material at the most convenient moment of their etymological discourse. For Wartburg, it is upon the base of the complete material, given in prime position in the article, that the analysis is constructed; the analysis being placed in the historical commentary. However, the manner of presentation itself constitutes a position on the etymology: clearly, and unavoidably, by the choice of the units which are going to be part of the article; less obviously perhaps by the organization of the article, which is by itself a discourse on the structuration of the descent from the etymon to the lexical units.

Thus the organization of the article gives the user the possibility of forming his/her own opinion: he/she can check the author's reasoning, since the whole material is made available. He/she can even check things in detail, since it is possible to return to all the sources, which are explicitly referred to (it should be pointed out, though, that, in reality, this work can be quite difficult to undertake). As we see in the article *mundus*, the final historical commentary only sketches the main facts, and it is the very setting out of the material which delineates the more detailed history.

I have underlined, in the article in the appendix, a few particular cases to show it. Col. 220a: *entremonde* appears two times in different subdivisions, since the first one is a creation on the basis of fr. *monde*, and the second one a borrowing to latin, hence two lexical units completely separate from an historical point of view. The same way, *l'autre monde*, col. 218a, and *l'autre monde*, col. 219b, are not the same locution, with an extension of meaning, but two independent creations on the basis of completely different meanings of *monde*. Similar comments could be made about *mondial*, 220a, a 16th Century word, and *mondial*, 219a, a 20th Century one, which is not based on the former one.

So can we characterize the FEW etymology as:

- concise
- not very verbalized
- not very explicit
- inherently understated

However, this has changed with current practice: we now set out explicitly and systematically the various affixes, which means that the number of footnotes is growing; and likewise for the

commentaries. Furthermore, the structure of the articles of the FEW is very flexible: the form of the articles is comprehensively adapted to the particular configuration of each etymological problem; the fundamental structure can be organized by morphological or semantic criteria, depending on the particular situation or the clarity of the resulting exposition. This flexibility and economy of the exposition, and hence – since it is the same thing – of the etymological discourse, of the etymologization, are a great virtue of the dictionary as it exists. So it finds it itself in the position of being able to express nuanced views on matters which, by their sheer size, would otherwise seem better suited to a more schematic treatment. However, these virtues of the paper version are, when digitization takes over, very serious impediments indeed.

2. The first digitization project looked at the possibility of creating a Document Type Definition in an xml-scheme. But this model, already difficult to implement in any polydialectal dictionary, proved quite clearly to be unworkable. The thought behind the digitization undertaking took a lot of time, since the constitution of a model for the digitization was the object of a PhD thesis, defended last June at the universities of Liège and Nancy by Pascale Renders. The model she designed and developed, with the collaboration of a computer scientist, relies fundamentally on the basic idea of making the computer programme structure comply to the structure of the dictionary itself, instead of imposing a different one. It consists, then, of an interpretation of the structure of the FEW's etymological discourse, rather than an interpretation of the etymologised units. The first difficulty is for the computer to understand a great deal of information concealed by the typically elliptic way of writing used by the FEW. In an FEW article, any information (the language, the form, the meaning, the grammatical category) has not to be given again if it remains the same, so when you are looking for a particular lexical unit, you can have to go up for a while in the article to find the complete information.

The simple sequencing of material is also significant for the etymologisation: 219a: it is the order of sequence which leads us to understand that *mondial* is the basis for derivation *mondialement*, then *mondialist* and *mondialisme*, and it is this order, too, which allows us to see the essential link between the last two. But the FEW does not make any difference between the relationship of the last two forms (whose formation is certainly interrelated) and the relationship of *mondialement* and *mondialisme*, which both depend on *mondial*. This discrimination is of too little importance, and the discourse content itself with presenting the material in such a way that the etymological analysis is made easy for the reader.

Similarly, the FEW must necessarily confront itself with contradictions in the classification: the dictionary distinguishes between the attestations of the lexical unit alone and the locutions in which it is found. But this distinction contradicts, in the § 3.a, the semantic classification, and, therefore, the different meanings are mixed in both alineas. This allows us to keep the simple lexical unit in the canonical geographical order of presentation. Nevertheless, Wartburg chose to allow mixing, in this enumeration, of the substantive and diverse cases of plurals accompanied by an article or a possessive. Thus he chose to present together some facts he could have chosen to keep apart: as they are near enough on the page, the reader can keep them simultaneously present in his/her mind, and can thus have an overview of all of the significant data. We can consider at the same time *ton monde* “your parents” and *mes mondes* “mes parents”, and also data for which we cannot say if the source discriminated them clearly from the general meaning “people in general”.

In such cases, the way of exposing data is hard to reconcile with the strict hierarchy of the complete sequencing that is implied in the tagging of every single unit for the purposes of the computer programme. And those cases are the absolute normality in an FEW article, where the alternation of numbered alineas can easily be rationalised and expressed with simple words, but where the balance of sections discriminated through dashes, dots, semicolons and commas can ask for much longer explanations. This poses a serious difficulty for the digitization.

Another difficulty is given by the presence of errors or inconsistencies, the different ways of presenting similar data, depending on the diversity of the authors and the long time of the writing of the dictionary. Indeed, it is quite difficult for a computer to cater for these. On the basis of Renders' thesis, there is actually a project of digitalizing the FEW, though at the present time we do not dispose of the financial means necessary to complete it.

3. There is something paradoxical in the FEW's etymologisation: the etymology, as we have seen, is largely given by the position of the analysed unit in the dictionary and in the article, but one has to know—or to guess, rather—this position in order to find the unit. In a sense, we could say that one only finds in the FEW a confirmation of the etymological hypothesis one set out to elaborate. Obviously, this is a rather simplistic account, since it neglects the commentaries, the notes, and the set of lexical units which are placed around the analysed one, but there is still a problem for searching within the FEW. You can see clearly that this searching problem can find a precious help, a help otherwise absent, in the computerization of the dictionary, and will find it. There is anyhow a contradiction as far as searching for isolated units is concerned since the dictionary never treats them in an isolated manner. In other words, if we are only looking for the FEW's article where to place a unit, and we think we have found something interesting, we are perhaps making the same mistake pointed out by Meillet when he said "the question vulgarly put to linguists 'What is the origin of a word?' has no precise meaning".

The reasons for digitalising the FEW are the easy searches for units, and the carrying out of searches using criteria that are not possible to use with the printed version. However, the fulfillment of these purposes includes some risks, and potentially the cutting of some corners, especially the temptation of renouncing reading. I have tried to show that the understanding of the dictionary relies on the reading of the whole article, or at least of the context of the form. However, with automatic searching or searching for large series of units, there is a risk, that the need for reading could be removed, the result of which would be that the etymological discourse is discarded in favour of the few items of information which have been included in the tagging. The digitization would lead directly to a change in the manner of reading the dictionary, but also, more importantly, to the loss of the etymological discourse elaborated by its authors, which would be a disaster.

But there is another problem for users of the digitized FEW to be, and that is the belief that he/she actually has found the lexical unit he/she was searching for. We are again dealing with the problems to do with the dictionary's inherent errors. The double etymologies, when they are erroneous, often reveal events in the history of lexicon: for example, attraction by another family or popular etymology. And in all the cases where the classification is correct (which is the vast majority), some forms may be classified in another article, for reasons which are made clear only by reading the second one. The sense of doubt experienced by the FEW's

user, if he/she really found the correct classification, is very useful, and there is some risk that automatic searching could lead to a false sense of confidence, even in the hypothesis that the computerization itself would not create any mistake. The act of reading, which implies the understanding of the dictionary, leads by itself to the critic assessment of the way units have been dealt with by the FEW.

4. I will turn now to a matter I discussed, a few months ago, in Aberystwyth: the problem of the corrections made to the dictionary. The FEW, as can be expected of such a big work, spanning 90 years, is full of mistakes and contradictions. But as the FEW is the reference for the lexicon of French and the Gallo-romance languages, every new advance is only validated by comparison with the status *quaestioni* present in the dictionary. Then the FEW is not only a discourse upon the history of words, but also a repertory of dated and localised forms, and a repertory of the bibliography of those forms and their relations. This is a function it has to assume, as long as nobody thinks of being able to write another repertory to replace it. It is, then, absolutely necessary to correct and complete the FEW, to keep up its place as a useful authority. This practice has been going on since the beginning, with the help of errata at the end of each volume, for example, but the digitization will allow us to use much more convenient tools for this purpose, and Pascale Renders, the author of the digitization project, has suggested making a place for the introduction of corrections in the digitized FEW, signaling them by a versioning system.

We have seen that there is an essential link between the represented material and the etymological discourse. We can then understand that any modification in the material will potentially lead to a modification of the etymological discourse. So we are, when we want to use the FEW as an authority for the history of French lexicon, in a paradoxical situation: any action aiming to strengthen this authority (new dating, correction of a form, addition of new forms) can result in undermining the dictionary. It is important to insist on the difficulty of introducing “*addenda et corrigenda*” in the FEW: to correct a date means maybe reviewing the entire chronological scenario of some part of the article; to correct a meaning could lead to a change in the classification of the unit, the concatenation of the meanings, could maybe lead to the discovery that a lexico-semantic unit is a regionalism, and this could lead to contest a derivation which is active only elsewhere, and then to change the etymology, etc. Any correction of this type can lead to the loss of the historical continuity of the attestations.

Adding a new form to the FEW amounts to a chance modification. And to integrate it in a correct way, by controlling which modifications this introduction induces means practically to rewrite the whole article. There is no simple way of correcting or adding to the FEW.

Thanks to the immensity of his collecting work, his admirable knowing of Romance studies and his personal merit, Walther von Wartburg created a work fundamentally coherent and correct. We do actually have quite a good idea of the FEW’s contradictions and failures, but they do not seem to modify the vastness and coherence of the overall conception. Introducing random modifications in this conception, even perfectly correct at a small scale, would destroy the unity of each article, and of the FEW as a whole. There would be no benefit for historical linguistics if there was, beside the FEW, a new FEW-version B. So, are we in a position of not being able to make corrections necessary to anyone’s eyes? Is it really impossible for the FEW digitization to allow what is allowed by every electronic dictionary?

The situation is not entirely helpless. We can find some elements of a solution, I believe, by considering that the FEW is not only an etymological discourse, but also a repertory of forms. As a repertory of the entire Gallo-romance lexicon, the FEW presents the best possible basis for classification, even when we consider its imperfections. If we cannot modify it easily, we can use it as a pure frame, to give easy access to a great mass of knowledge of the lexicon, published elsewhere. I propose then to use the digitized FEW as a repertory giving access to the suggestion of additions or corrections. It will not be a modified form of the dictionary, but the simple use of its structure. So we will not have to validate or examine the correctness of the modifications: there will only be simple references to published suggestions, and the responsibility of the correction will lie with the author and the journal where it is published. What are the consequences of such a modification for the FEW? The most important one is that there will not be any new versions of the FEW. The second one is that criticism of the dictionary will get easier, whilst its use becomes harder, since the use will lead naturally to the criticism and the user will be guided in reviewing some parts of the etymological reasoning.

The necessity of determining by oneself, on the basis of addenda and corrigenda, if the etymological discourse of the FEW is still valid, is naturally a great problem. But maybe it is not such a serious problem: by choosing not to consider, as we can be tempted to, the etymological discourse of the FEW as an evidence given by the facts, we will see it more clearly as immanent result of the so powerful conception of Wartburg.

The elaboration of the dictionary's general digitization is not an impediment to the completion of partial projects, related to it, such as:

- a programme of automatic cartography (soon to be achieved);
- the realisation of an onomasiological index (partially available on our site, being progressively completed);
- the possibility of searching automatically in the list of etymons;
- a (very) partial tagging of suffixes in the dictionaries

sommes bien mises, élégantes" (seit 1798). - Nfr. avoir son monde „savoir bien la manière de vivre dans la société" (seit 1612, Régnier); avoir du monde (seit 1671, Sév); n'être plus du monde „ne plus fréquenter la société" (Ac 1694-1878); homme, femme du monde „homme, femme qui vivent dans la meilleure société" (seit 1660, Reiz).

Nfr. demi-monde „ensemble des prostituées" (1804, Bern). Constant, Journal intime 1851 ?), „société des femmes de mœurs légères" (seit 1855 ?); quart de monde „société tout à fait dévergondée, mais qui affecte certaines prétentions à l'élégance" (Laroch 1861-Lar 1903). - Nfr. demi-monde „société ambiguë des aristocrates" (1885, Maup, Bossion 42). 4. Fr. monde „vie du siècle" (seit Ruteb, s. Doch). Abl. - Mfr. mondialis adj. „du monde (par oppos. à spirituel, divin)" (ca. 1608-1646); mondialisé f. „mondialisé" (ca. 1608). III. Mfr. centromonde m. „ce qui est entre les mondes" (Amyot-1608).

It. mundus „das weltall, das universum" lebt weiter in it. mondo, legid. mundi, obeng. mundi, fil. moni, lat. mon sowie im gallrom. (oben I 1). Seit dem 1. jh. vor Chr. wird mundus auch mit Bezug auf die Erde allein verwendet (so bei Horaz), daher 2. Daraus entsteht weiterhin die bed. „Gesamtheit der Menschen", vgl. Horaz quem mundus et superi iuvant (3). Die christlichen Schriftsteller stellen mundus in Gegensatz zum geistigen, religiösen Leben, so in der Vulgata Joh. 18, 36 regnum meum non est de hoc mundo; Hieronymus Ep. 38, 2 mundus mori. Diese bed. ist in 4 bezeugt - Seit dem 12. jh. wird mon allmählich durch monde ersetzt, das neu aus dem le entlehnt ist. Dieser Ersatz ist nötig geworden wegen der Mehrdeutigkeit von mon (s. mons, monstrum). Im apr. scheint er später erfolgt zu sein, als im fr. Die semantische Entfaltung entspricht denjenigen von I. Bei 3 und 4 scheint der Ersatz von mon durch monde etwas später eingesetzt zu haben als bei 1 und 2. Es gibt auch Texte, die mon und monde nebeneinander brauchen, so ContPere, BestG (mont 1074, monde 1075). Bei 3 ist noch eine semantische Sonderentwicklung eingetreten: gegen Ende des 16. jhs. bildet monde aus der bed. „Orte" her aus noch die von „gesellschaftliches Zusammensein der Menschen" (b. gegenüber 2). Der nach der Bedeutung der religionskritischen einsetzende drang nach gemeinsamem Genuss des Lebens hat wohl viel dazu beigetragen. Im 16. jh. wird vorübergehend auch lt. intermundici „zwischenräume zwischen den weltl." entlehnt (III). - Berger: ML 5749.

1) Vielleicht auch schon 1136 bei Wac, doch kann es sich auch um eine etwas spätere Handschrift handeln.
2) Übersetzung von gr. μυνηστικός „kleine welt, mensch".
3) So un monde d'or Chastell, un monde de vices Calvin, un monde d'effrayes Didot. Gebrauchen in un monde d'illusions, d'idées.

4) Die Tatsache, dass monde ein Kollektivum ist, führt dazu, dass sich die Vorstellung einer mehrzahl damit verbindet, woraus dann verschiedentlich sich wieder der Gebrauch von monde für eine einzelne Person ergibt. Aus diesem hin und her zwischen Kollektivum, sing. und plural ergibt sich manchenorts ein gemischter Gebrauch, indem das Subst. im sing. mit dem Verbum im plural verbunden wird, so pop. que de monde qui vont par là „que de monde qui va par là" Desgr 1821, bman. i mot vö „les gens vont". Ähnlich hman. Chastell. Couson. So auch pop. tout le monde disent Platt 1836.

5) Hierher wahrscheinlich auch Anney mondi Thénas mondi, Pois mondu, die aber nur mit „monde" dekliniert sind.

6) Die bed. „le monde entier", die tout le monde seit dem 12. jh. auch hatte, weicht seit dem 18. jh. vor der bed. „die es im Rahmen der bed. „gens" hat, Andersson hält tout le monde, tous les hommes für eine Neuschöpfung des französischen. Doch zeigt Bambeck Z 74, 36b, dass schon im spät. tous mondi in beiden bed. gebraucht wird. Vgl. Valgata Evangelium in toto mundo „... dans le monde entier", neben Evangelium 12, 19: ecce mundi totus post eum abiit „tout le monde court après lui".

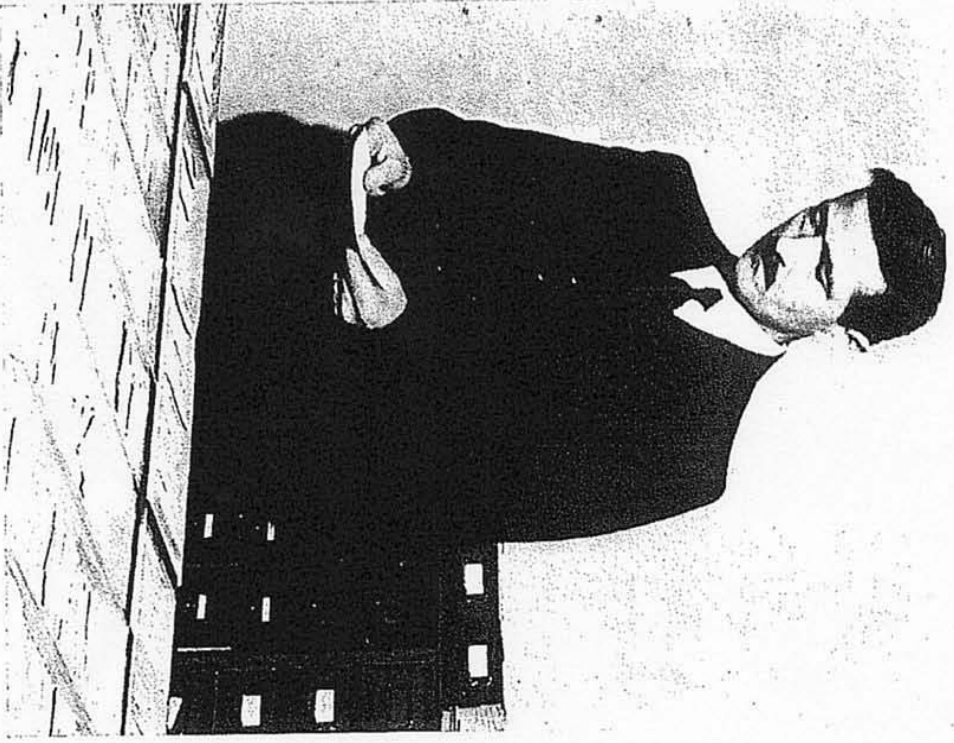
7) Ayant en occasion d'avoir quelques détails intimes de l'existence des filles du demi-monde, je me suis convaincu que le roman de Justine n'est point une exagération de la vie humaine. Quelle espèce que la nôtre! Justine ist ein 1791 erscheinender roman des Marquis de Sade.

8) Das Wort ist in dieser bed. zuerst von Dumas fils gebraucht worden, der 1855 eine Komödie Le Demi-monde über die bühne gehen liess; darin werden Frauen dargestellt, die sich hingeeben haben, aber dem Schein noch beharren möchten. Die bed. des Wortes hat sich dann aber rasch nach der nance entwickelt, die es heute hat (so schon Delv 1867). - Darnach auch demi-mondaine, s. MUNDANS.

mündus sauer.

I. Apr. mon, -da adj. „pur, net" (13-14. jh.). - Abl. Cr. tot mundamini „tout doucement", Crampone mondamin, Riviere „un tant soit peu". Apr. der-mondar v. a. „sonneler" (ca. 1280).

II. Afr. mundi adj. „pur, sans souillure, vertueux" (Wace 13. jh.). Afr. mfr. monde (Christien-Oud 1680, Gdf: TL; Jeh Bouche; HonnAde; MeHee 312; PassPfr: Guilmach. Comm; ColenAde; Hu; Saliel; FritSaliel), mund Moan, Apr monde „net" M. SallSP, monde „pur" RLR 26, 60. Afr. monde de „dépouillé, privé de" (13. jh., mfr. id. Desch. Bures monde, -o „faible" Castet 52; mfr. monde „dont on peut se nourrir et qu'on peut offrir en sacrifice, selon la loi juive (animal)" (ca. 1850), mfr. id. Pur 1680-Lar 1948). - Abl. Afr. mondement adv. „d'une façon pure, vertueuse (au moral)" (13. jh.), mfr. „d'une façon propre, nette" Huls 1696; for. tout mondement „très peu". Mfr. mondan adj. „non altéré (de la viande); bien fourni (de légumes)" (Att- ca. 1380). Mfr. mondin m. „homme propre,



Basel
1939-1971