Geert De Wilde, 2011 ‘Re-Considering the Semantic Labels of the Anglo-Norman Dictionary;. in David Trotter (ed.), *Present and Future Research in Anglo-Norman: Aberystwyth Colloquium, July 2011* (Aberystwyth: The Anglo-Norman Online Hub, 2012), 143-50.

ANGLO-NORMAN DICTIONARY

Re-considering the semantic labels of the Anglo-Norman Dictionary

Geert De Wilde, AND, Aberystwyth

From the first fascicle of its first edition, published in 1977, the AND has made use of (what appear to be) semantic usage labels for certain entries.[1] These labels, which mark the use of technical terms, are set apart from the actual definition by means of round brackets and the absence of italics. To give two random examples taken from its first few pages: the entry **abatre**[1] lists as its final sense '(law) *to abate, put an end to*', to distinguish this one from its non-legal senses, and **ache**[1] is glossed '(bot.) *wild celery*'. The brief 'Introductory Note' of the first fascicle (AND1, vii) does not comment on these labels or their purpose, and no comprehensive list of them is provided anywhere. A considerable portion, however, appears in the section 'abbreviations' (AND1, viii), but, evidently, only those that *are* abbreviations. Consequently, 'bot.' is listed as 'in botany', while tags like 'law', 'material' or 'local' are absent. None of the later fascicles, including William Rothwell's more extensive 'General Preface' found in fascicle 7, return to the subject. The upshot is that throughout the dictionary's first edition, the use of these technical usage labels was considered self-explanatory, and no indication was given either of any editorial intent or of their range.

In 2005 David Trotter wrote a new 'Reader's Guide' for the second edition of the AND, which was published in Vol. I (A-C) as well as online.[2] Here he gives a detailed outline of the contents and structure of a dictionary entry: "Articles indicate first the part of speech [...], then supply a gloss (italicised) and a quotation or quotations (in roman) illustrating that sense" (xxiv). In other words, the existence of any label, be it a language tag or a usage label (in brackets and not italicized), between the part of speech and the gloss, is once again not even acknowledged. Thus, it appears that AND2 has completely taken over AND1's non-policy on the use of (semantic) labels.

Despite appearances, things *have* been changing: to begin with, the creation of an on-line version of AND2, which introduced the use of a more consistent and uniform format to all articles, also brought along a reconsideration of the status of the semantic label. When the Word-based version of the second edition of A to E was converted into an XML version, the decision was made to distinguish these labels from the actual definition by tagging them separately. Thus, the 'usage tag' was introduced, which precedes the 'translation tag', rendering one of the above examples as <usage type"bot."/> <trans>wild celery</trans>. Most significantly, this XML 'usage tag' was envisaged not as a free-form section but as a 'pick-list', which means that only a defined (albeit expandable) set of labels became available to be used. The way this set was created, was by a straightforward (and in the first instance automatic) consolidation of all the semantic labels that already appear in the dictionary. As such, it uses AND2's list of abbreviations as a basis (removing some but not all of the non-semantic material), and added all other bracketed information found in that position (but excluding the language labels). Not surprisingly, this brought to light several inconsistencies,

---

[1] For the first edition (AND1), see *Anglo-Norman Dictionary*, ed. by William Rothwell *et al.*, 7 vols, Publications of the Modern Humanities Research Association 8 (London: Modern Humanities Research Association, 1977-92).

[2] The second edition of the Anglo-Norman Dictionary (AND2), currently complete from A to L (with M forthcoming in 2012), is published digitally and available online: www.anglo-norman.net. A printed version was prepared for the section A-E: *Anglo-Norman Dictionary: Second Edition*, ed. by William Rothwell and others, 2 vols, Publications of the Modern Humanities Research Association 17 (London: Maney Publishing for the Modern Humanities Research Association, 2005). For the reader's guide, see Vol. 1, xxiv-xxvii and http://www.anglo-norman.net/sitedocs/main-intro.html#sec3.

not only with numerous presentation variants of the same label, (e.g. 'anat', 'anat.' and 'anatomy'), but also with the rather liberal use of the bracketed label for all kinds of information, such as 'as title of book', 'of children, hawks, hounds', or 'of fracture of skull', that do not indicate a general semantic field but merely provide further information on the context. Phrases like these had to be moved, at this point, inside the definition or, in some cases, were tagged in a different way (as a more general 'note'). The resulting XML pick-list of semantic labels (which was the first 'complete' list of that kind), after ironing out any overlaps, revealed a considerable number of genuine 'new' labels that were never added to the 'abbreviations' section, such as 'acad.' (academic), 'culin.' (culinary), 'myth.' (mythology), etc. From the revision of F onwards, the editors of AND2 have confined themselves to this list for the writing of all entries, and it is no longer possible to create ad hoc semantic labels. At this stage, any new label that is deemed necessary, first has to be added, in a separate exercise, to the XML pick-list.

While this implementation first acknowledged the truly separate status of the semantic label, it also exposed the problematic nature of that status – which becomes most apparent when, in the case of the online AND with its range of different search facilities, the question is posed whether it is possible to perform a search by semantic label. For example, the *Lexis of Cloth and Clothing Project* in Manchester has asked if we have a separate label for clothing (the answer was, unfortunately, no),[3] and a project like the *Dictionnaire du Français Scientifique Médiéval* in Paris might want to call up all articles which use labels such as arithm. (arithmetic), archit. (architecture), hort. (horticulture), agr. (agriculture), etc.[4] Similarly, labels like 'her.' or 'mus.' would make it possible for a user to bring together all heraldic or musical terms, and as such to do research on a particular (semantic) field of the language. For AND-editors it would offer the possibility of analysing the language by semantic group, and allow them to identify possible lacunae or under-represented areas. Although the separate tagging in the underlying XML currently allows for the possibility of introducing such a search option, we have, so far, decided not to make this available to the general public yet. There are two main reasons for this.

Firstly, as a direct result of the lack of any editorial statement on this, the use of semantic labels has, unfortunately, been treated at times almost erratically and as something entirely optional in the writing of a dictionary entry. The bracketed bit of information was only added, it seems, when a given editor believed it would clarify or improve his or her definition. For example, the entry **aristologie** is glossed 'aristolochia' and benefits greatly from the tag 'bot.' to clarify that what we have here is actually a plant. In contrast, the article **arbre** for its main and generic sense 'tree' did not seem to require the botanical label. Nevertheless, in the same article, there is a label for the locution *arbre de basme* ('balm-tree') but there is none for *arbre pomer* ('fruit-tree'). In comparison, the entirely generic entry **plante**[1] was provided with a 'bot.' label. In a similarly inconsistent way, **lance** (glossed 'spear, lance') has the label 'mil.' for military, whereas **espee** ('sword') comes without such a label. Furthermore, some new labels were introduced over the years, only to be ignored or forgotten about in later entries. For example, the label 'currency' was introduced for the article **angevin**, and only reappears for **livre**[2]. Other entries for currencies use the label 'coin', like for example **ferthing** or **marc**, whereas quite a few do not use any label at all. Add to this the fact that different editors have interpreted the implications or even meanings of semantic labels differently (the most striking example being the tag 'iron.' appearing not only with reference to the metal but also to signal the 'ironic use' of a word), and the result is an at times highly inconsistent distribution of the existing labels. As a result, any search for a particular semantic label would not offer any form

[3] http://lexisproject.arts.manchester.ac.uk/research/index.html
[4] http://crealsciences.univ-paris13.fr/

of completeness or reliability. A well-used label, like for example 'law' appears in more than four thousand entries, whereas a 'forgotten' label like 'metal' appears in just four.

A second reason for holding back the search by semantic label facility is the incomplete and, unfortunately, still disorganized nature of the current defined list of semantic labels. As described earlier, there never was a point in the history of the dictionary, when editors came together and decided which semantic labels would be required to cover all possible angles and fields expected in a medieval language. Instead, the only list of existing labels was produced in house, and is merely a collection of those created more or less ad hoc over a period of several decades. This has resulted in a great number of oddities and anomalies, of which I will quickly mention a few.

Firstly, different tags have sometimes been used to refer to the same semantic field, or if there are subtle distinctions, those do not seem to have been adopted in their usage. For example, we have the two labels 'arch.' and 'archit.' to refer to 'architecture', which, arguably, overlap with the tag 'build.'. Similarly, the labels 'topon.' (toponomy) and 'geog.' (geography) have at times been used indiscriminately to refer to place-names. The label 'geog.' (geography) is also used for features of the landscape, which overlaps with the label 'topog.' (topography). And then we have competing labels such as 'math.' (mathematical) vs. 'arithm.' (arithmetic), or 'mar.' (maritime) vs. 'nav.' (naval), where different editors just seem to have preferred different nomenclature.

Secondly, the dictionary uses some labels that are non-medieval. One example of this is the distinction between 'astrol.' (astrology) (appearing 21 times) and 'astron.' (astronomy) (appearing 34 times). It is clear that trying to differentiate these fields would be a complicated task which is anachronistic to the medieval way of thinking. Instances like these should be avoided in the dictionary.

Thirdly, the defined set of usage tags still contains a considerable number of non-semantic labels. For example, two heavily used tags are 'fig.' (figurative) and 'coll.' (collective) – these labels qualify the usage of a word one way or another, but they do not make any difference in terms of its semantic field. Similarly, tags like 'iron.' (ironic) 'pej.' (pejorative), and 'vulgar', or even 'imprecation' and 'exclamation' belong to a different level of language interpretation and would therefore best be detached from this group by using a different tag – in this case one that merely signals the register.

Fourthly, we have a number of usage tags that have been used only a few times throughout the dictionary, so that their semantic width is not obvious. For example, 'hist.' probably stands for 'historical', but it is not clear what in a dictionary of a medieval language stands out as more 'historical'. It appears twelve times, in articles as diverse as ***baston*** ('a warden of the Fleet prison who carried a red staff as a symbol of office'), **ju**[1] (attached to the locution *ju d'Olimpiades*, 'Olympic games') and **merchet**[1] ('fine paid to overlord for permission to give one's daughter in marriage'). With a clarification of this lacking, current editors are often hesitant to use a tag like this again.

Lastly (and most importantly), there are several semantic areas that have been covered incompletely or hardly at all. We currently have tags for words for fishes ('ich.') and birds ('orn.), and it even turned out that, for reasons thus far unexplained, we have two tags for horses: 'horse' and 'horses'. Other animal-words in the AND do not come with a specific semantic label, although eighty-five have been tagged as 'zool.' (zoological) – ranging from dogs, to seals, hedgehogs, lizards and even hornets. The labels for fish- and bird-names are strictly speaking sub-categories of the 'zool.' one, so if we distinguish those, are there any other animals or groups of animals we have to separate? Do we need further semantic labels such as 'reptiles' or 'insects' or 'domestic animals' and so forth? As a second example of the

partial representation of certain semantic sub-groups, we have the general tag 'games and sports', as well as the more specific 'chess' (probably prompted by the inclusion of a treatise on chess in the List of Texts) and somewhat surprisingly 'wrestling'. For semantic fields which are currently not covered at all, I could mention oenology (or wine-making), lapidary (names of stones), philosophy, units of measure, clothing, art, non-Christian religions, etc.

With the AND currently presenting itself as primarily an online dictionary, to which the wide range of search options forms an integral functional part, the editors now face the challenge of resolving the abovementioned state of affairs, and of turning the semantic label into a more reliable and searchable feature of the dictionary. Evidently, because of the sheer scope of such a project (which would involve returning to every single article in the dictionary to re-assess and expand its semantic tags), it is essential to 'get it right' this time, and to create a classification system which is both comprehensive and transparent. The crucial starting-point would be the putting together of a full list of all the semantic labels that are going to be required, which not only fills in the areas that are currently missing but also removes those usage tags that are not semantic (ironic, figurative, etc.).

In order to achieve this, other dictionaries that have already produced comparable semantic lists should be used as guidance. A first possible model is the list of 'disciplines' of the online *Trésor de la Langue Française*.[5] This list, which forms part of the TLFi's 'Recherche assistée', creates twenty-one general 'centres d'intérêt', such as 'Arts et spectacles', 'Enseignement', 'Médicine, santé' and 'Sciences occultes', with numerous sub-categories for each field. For example, 'Science occultes' subdivides into 'alchimie', 'astrologie', 'chiromancie', 'occultisme', etc. Evidently, the TLFi's list as it stands goes far beyond the medieval range of meanings, and could therefore not straightforwardly be applied to the AND.

In a very similar way, the Oxford English Dictionary allows one to specify certain 'categories' in its 'advanced search', which consist of twenty-one (different) 'subjects', such as 'agriculture and horticulture', 'heraldry', 'law' and 'military', divided into sub-groups as well.[6] It distinguishes these as a set from a different search 'layer' entitled 'usage', which includes categories such as 'derogatory', 'euphemistic' etc. As already mentioned, a similar differentiation would be required for the AND usage tags. In the OED, a user can browse this 'subject' list and by clicking on a particular category retrieve all relevant senses in the dictionary. These senses can then be narrowed down into even more specific sub-categories. Interestingly, in the articles themselves, the subject-category is sometimes, but certainly not always, explicitly stated in the definition, which suggests that this search facility by semantic category is independent from the (visible) contents of the articles. Thus, by making the semantic label(s) invisible in the entry or sense, the OED avoids cluttering the definition, while still allowing the search facility to function accurately.

In 2010, the OED also added the 'Historical Thesaurus' as a search tool to their website, which provides a much more detailed and taxonomic classification of most of its senses.[7] This independently created thesaurus not only constructs a semantic index of the entire English language, but also allows the OED to function fully as an onomasiological tool. While this goes far beyond the function of the semantic label as envisaged for the AND, it should be kept in mind that the OED's Historical Thesaurus has the potential of serving as a central point of reference for any semantic labels, and could therefore possibly be linked to any system the AND might create.

---

[5] See http://www.atilf.fr/tlfi

[6] See http://www.oed.com/browsecategory

[7] See http://www.oed.com/thesaurus.

Volumes 21 to 23 of the *Französisches Etymologisches Wörterbuch*, which deal with *Materialien unbekannten oder unsicheren Ursprungs*, provide an example of a truly onomasiological concept-orientated presentation of a dictionary, whereby the entries are organised not in alphabetical order but by their meaning as part of a universal semantic classification structure.[8] The basis of this is a *Begriffssystem*, originally developed by Wartburg in collaboration with Rudolf Hallig, which is similar to what is achieved by the Historical Thesaurus of the OED.[9] For example, the first section 'L'univers' has the categories 'Le ciel et l'atmosphère', 'La terre', 'Les plantes' etc., with the latter subdividing in 'La vie végétale en général', 'Les arbres', 'Les arbrisseaux et plantes à baies', 'Les plantes alimentaires' etc. Each section then makes more subtle differentiations until it arrives at the actual senses. As with the Historical Thesaurus, such a semantic taxonomy makes it possible not only to reveal areas that are relevant for the mapping out of the AND's usage tags, but also to rely on a central point of reference that locates those labels within a general semantic framework, which would then enable the AND to link more directly to other dictionaries using the same framework.[10]

A fifth and last example can be found in the *Complément* to the FEW, where a much more general list of twenty 'domaines spécialisés' which refer to the areas dealt with by specialised dictionaries that the FEW uses as sources.[11] These distinguish broad semantic domains such as 'armée', 'arts', 'botanique', or 'chasse', etc., and form a collection which is much more similar to the one currently used by the AND. Nevertheless, it already has categories which are absent from the AND, such as 'commerce' or 'métiers'.

These five examples clearly demonstrate how much of a difference there can be in the range of a semantic categorisation, and, from the outset, a decision would have to be made on how detailed and precise the AND semantic list should become – on how far down the road of an underlying onomasiological dictionary it should go.

I would like to highlight two practical requirements, or caveats, to take into account with a view to the formation of such a semantic list. My first caveat is that this list, as mentioned before, should have maximum transparency, both for the users and the editors. One way to achieve this is by grouping several semantic labels together into more general fields, similar to what has been done in the TLFi or the OED. For example, 'anat.' (anatomy), 'zool.' (zoology), bot. (botany) could all go together under a heading 'biology and nature'. In the same way, 'games and sport' could be a general heading for labels such as 'chess', 'wrestling' and others. In this way, the AND would construct a minimal semantic tree or hierarchy that would not go as far as the Hallig / Wartburg or OED's Historical Thesaurus taxonomies, but that would make it easier simply to maintain an overview of the different categories. Another option (which none of the above-mentioned dictionaries seem to use) to increase transparency would be to add, at this early stage, definitions or editorial statements which clarify and specify these semantic labels. For one thing, this would enable continuity between different

---

[8] Walter v. Wartburg *et al.*, *Französisches Etymologisches Wörterbuch: Eine Darstellung des galloromanischen Sprachschatzes*, vols 21-23, *Materialien unbekannten oder unsicheren Ursprungs* (Basel: Zbinden, 1965-97).

[9] Rudolf Hallig and Walter v. Wartburg, *Begriffssystem als Grundlage für die Lexikographie; Versuch eines Ordnungsschemas* (Berlin: Akademie-Verlag, 1952).

[10] The *Analyse et Traitement Informatique de la Langue Française* laboratory (ATILF), http://www.atilf.fr/, is currently preparing a digital version of the Hallig/Wartburg *Begriffssystem*, which ideally could serve as the foundation of an onomasiologically-based system to link different dictionaries.

[11] Chauveau, Jean-Paul / Greub, Yan / Seidl, Christian, *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes von Walther v. Wartburg, Complément*, 3rd edition, Bibliothèque de Linguistique Romane, Hors Série 1 (Strasbourg: Éditions de Linguistique et de Philologie, 2010), 353-55.

editors and over longer periods of time, avoiding confusion between tags such as, for example, currently 'theol.' (theological), 'eccl.' (ecclesiastical) and 'bibl.' (Biblical).

The second requirement is that to a certain extent, multiple labelling will have to be allowed. Currently the XML pick-list has a small number of labels that have a two-fold reference, such as 'eccl. and law', 'mil. and nav.' or 'mus. and fig.'. These are not only awkward in that allowance would have to be made for a great number of possible combinations (for example, if we have 'eccl. and law', what about 'nav. and law', 'agr. and law', 'forestry and law' etc.), but they would almost certainly create complications for search engines. A more straightforward solution would be that a sense could have more than one semantic label attached (which is currently not the case in the online AND). For example, an entry like **barge** would need the label 'nav.' (naval) that applies to all its senses, and with extra labels 'mil.' (military) for 'war vessel', 'merc.' (mercantile) as well as 'measure' (unit of measure) for 'barge load', and 'her.' (heraldic) for 'depiction of a barge, galley'. For the time being, I leave open the question whether the multiple labels system should then also reflect the aforementioned semantic hierarchy. In other words: should, for example, all entries with an 'orn.' label (bird names) also, perhaps invisibly, have a 'zool.' (zoological) label as well as a 'biology and nature' label attached? It will, perhaps, be a matter of running a small-scale trial section in the dictionary to find out to what extent and in which ways such a presentation, which could, of course, be semi-automatically applied, would be workable and/or useful.

In conclusion, the reworking of the semantic labels has the potential to become a major editorial task that might even have to be set up, initially, as a separate project. It is therefore essential that the editorial team has a clear idea, from the outset, of where they want such a revision to lead and of what level of semantic differentiation users would expect. It is, however, already clear from similar dictionary projects that, given the nature of the online AND, a feature like this would substantially enhance the quality and the editorial consistency of both the dictionary and its search-facilities.

Appendix: semantic labels / usage tags currently available in AND2

acad.
accounting
agr.
anat.
arch.
archit.
arithm.
as an armorial bearing
astrol.
astron.
Bibl.
bot.
build.
chem.
chess
coin
coll.
comparative
culin.
currency
decoration
eccl.
eccl. law
excl.
exclam.
fig.
fishing
food
forestry
games and sports
geog.
gram.
her.
hist.
horse
horses
ich.
imprecation
iron.
law
law and mil.
letter
lit.
lit. and fig.
local
local and temporal
logic
mar.
material
math.
med.
med. and astrol.
med and fig.
merc.
metal
mil.
mil. and her.
mil. and nav.
mus.
mus. and fig.
myth.
nav.
orn.
painting
pej.
pharm.
politeness formula
prov.
temporal
textiles
theol.
title
topog.
topon.
ven.
vulgar
weather
wrestling
zool.

Bibliographical references

Chauveau, Jean-Paul / Greub, Yan / Seidl, Christian ([3]2010): *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes von Walther v. Wartburg, Complément* (Bibliothèque de Linguistique Romane, Hors Série 1). Strasbourg: Éditions de Linguistique et de Philologie.

Ducos, Joëlle *et al.* (2008-): *Dictionnaire du Français Scientifique Médiéval*. Paris. http://crealsciences.univ-paris13.fr/.

Hallig, Rudolf / Wartburg, Walter v. (1952): *Begriffssystem als Grundlage für die Lexikographie; Versuch eines Ordnungsschemas*. Berlin: Akademie-Verlag.

Owen-Crocker, Gale R. / Sylvester, Louise / Warr, Cordelia *et al.* (2006-): *The Lexis of Cloth and Clothing Project*. Manchester. http://lexisproject.arts.manchester.ac.uk/research/index.html

Rothwell, William / Stone, Louise / Reid, T.B.W. *et al.* (1977-92): *Anglo-Norman Dictionary*, Publications of the Modern Humanities Research Association 8 (7 vols). London: Modern Humanities Research Association (AND1).

—, Gregory, Stewart / Trotter, David *et al.* (2005): *Anglo-Norman Dictionary: Second Edition, A-E*, Publications of the Modern Humanities Research Association 17 (2 vols). London: Maney Publishing for the Modern Humanities Research Association. www.anglo-norman.net. (AND2).

Simpson, John / Weiner, Edmund *et al.*: *Oxford English Dictionary*. Oxford: Oxford University Press. http://www.oed.com. (OED).

Trésor de la langue française informatisé. http://www.atilf.fr/tlfi (TLF).

Wartburg, Walter v. *et al.* (1965-97): *Französisches Etymologisches Wörterbuch: Eine Darstellung des galloromanischen Sprachschatzes*, vols 21-23, *Materialien unbekannten oder unsicheren Ursprungs*. Basel: Zbinden (FEW).