Semantic Problems of Thesaurus Mapping

Martin Doerr
Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Crete, Greece
Email: martin@ics.forth.gr

Abstract
With networked information access to heterogeneous data sources, the problem of terminology provision and interoperability of controlled vocabulary schemes such as thesauri becomes increasingly urgent. Solutions are needed to improve the performance of full-text retrieval systems and to guide the design of controlled terminology schemes for use in structured data, including metadata. Thesauri are created in different languages, with different scope and points of view and at different levels of abstraction and detail, to accommodate access to a specific group of collections. In a wider search accessing distributed collections, the user would like to start with familiar terminology and let the system find out the correspondences to other terminologies in order to retrieve equivalent results from all addressed collections. This paper investigates possible semantic differences that may hinder the unambiguous mapping and transition from one thesaurus to another. It focusses on the differences of meaning of terms and their relations as intended by their creators for indexing and querying a specific collection, in contrast to methods investigating the statistical relevance of terms for objects in a collection. It develops a notion of optimal mapping, paying particular attention to the intellectual quality of mappings between terms from different vocabularies and to problems of polysemy. Proposals are made to limit the vagueness introduced by the transition from one vocabulary to another. The paper shows ways in which thesaurus creators can improve their methodology to meet the challenges of networked access to heterogeneous systems.

1 Introduction
Terminological resources are increasingly important for information retrieval in wide area networks, for retrieving documents by querying databases and metadata employing controlled vocabularies. In particular, thesauri which organize terms and associated concepts in the form of simple semantic networks become important tools for searching through the rapidly growing electronic information flood. There is growing interest in developing automated intermediaries to negotiate the differences between controlled vocabulary schemes so that a user can use a familiar set of terms to search collections using other vocabulary schemes.

The paper discusses the effect of thesaurus mapping on the vagueness in retrieval from a theoretical and logical point of view, separate from the effects of the relation of the thesaurus to the collection it addresses. Therefore, it makes ideal assumptions for the latter without going into any detail. It assumes that a certain collection is tightly connected to a terminology resource, which is given in the form of a thesaurus containing typed relations between terms and concepts, as defined by Foskett (1997). By "tight connection" we mean, in one case, that data values or classification terms are restricted to this thesaurus and consistently used. In that case, a query with some term should retrieve exactly the objects meant (this may actually require the expansion of a query term into its narrower terms). Let's refer to this case as the controlled vocabulary situation. Alternatively, the collection may use free text or free keywords. In this case, what we mean by "tight connection" is that a search-aid thesaurus exists that approximates the situation and consistently used. In that case, what we mean by "tight connection" is that a search-aid thesaurus exists that approximates the expert language used in the collection and the associated concepts. In this case, we assume that this thesaurus provides better results for that resource than any other. Let's refer to this case as the free text situation. Finally, we are not concerned with documents only, but with museum object descriptions and others as well. The term objects is used here to cover all collection objects, not just text documents.

1.1 Related Work
There is a vast literature and established practice about creating thesauri for the purpose of information description and retrieval (for example, publications of the Getty Research Institute, the British Arts and Humanities Data Service (AHDS), and various national standards). Thesauri are designed as an agreement and compromise on a set of shared terms for common concepts. Even though the agreement on term definitions is sought over large communities, thesauri even in the same domain differ significantly (Krause 2000), and there is limited interoperability between tools and digital resources employing different thesauri.

To resolve the incompatibility between different terminology resources, research has initially concentrated on attempts to unify thesauri by merging, e.g. Mili and Rada (1988), Mannino et al. (1988), Rada and Martin (1987). The Unified Medical Language System (UMLS) merges concepts from some 50 sources into a metathesaurus, which retains links to its original sources. It is probably the largest merging effort undertaken so far (Nelson 1999). Two problems have turned out to be the most difficult:

- First, differences in term semantics, semantics of hierarchical relations and term overlap can render the simple combination of concepts from two sources impossible. Resolutions to this problem are usually semiautomatic, see e.g. Constantinopoulou and Sintichakis (1997), and can become fairly expensive.
- Second, controlled vocabularies are often associated with a large installed base of systems using them, such that migration to a new set of terminology and relations may be virtually impossible; for example, with subject headings used by national libraries (Chan 2000, Landry 2000).

Recently the focus has been more on creating systems that provide a transition from one thesaurus or subject heading scheme to another. Two paradigms exist:

- thesaurus correlation (ISO9594 (1985), Getty Information Institute (1996), Doerr and Fundalik (1998a)), where substituting concepts are sought
- thesaurus federation (Kramer et al. 1997), where lead-in terms to related concepts are sought.

Nikola et al. (1999) propagated both approaches together. The German CARMEN project has begun to correlate different German thesauri used to index social science literature, using intellectual and statistical methods simultaneously. In the MACS project (Landry 2000), the Consortium of European National Libraries (CENL) is currently investigating the feasibility of correlating the subject headings of all European languages. When terminology in different natural languages is combined, the
result is often referred to as "multilingual thesaurus." In the TermIT project it was found that creating multilingual thesauri and combining different thesauri in the same language belongs to the same class of problem (supported by Krause 2000). More about “multilingual thesaurus” correlation is included in section 2.2.

The problem of combining subject headings and classification systems, such as Library of Congress Subject Headings (LCSH) and Dewey Decimal Classification (DDC), has been investigated by Chan (2000) and Vizine-Goetz (1998). Chan writes: "How to combine the salient features of a rich vocabulary like LCSH and the structured hierarchy found in classification schemes such as LCC and DDC to improve retrieval of networked resources remains a fertile field for research and exploration", and advocates the harmonization of vocabularies. Chan refers to the efforts of the Library of Congress to make the LCSH more useful for networked access by improving its faceted structure, the principles of term construction, and rigorous term relationships. In this paper, we argue for development in the same direction.

1.2 Intellectual versus Automatic Term Correlation

Term correlations may be intellectually created, as by MACS, the French Ministry of Culture (see Merimee), the HEREIN Project, and CARMEN (Krause 2000). Krause calls intellectually created term correlation lists “cross-concordances.” Alternatively, they may be created by statistical methods or even by neural networks, as in CARMEN and Vizine-Goetz (1998). Chen et al. (1996) use a concept space approach to create thesauri automatically and to traverse between two thesauri from different biological subdomains. They report a considerable increase of recall in a cross-domain search experiment, but differences between the links provided by the algorithm and those given by experts. Statistical and neural network methods do not easily allow interpretation of the intellectual nature of a given link, if at all. They are far cheaper, however, and can detect relations of which humans are unaware. The ultimate precision is usually low. As Chan (2000) puts it: "the tension between quality and quantity has never been keener.” As this paper deals with semantic problems, we do not consider statistical methods, even though we are convinced that the future lies in the coordinated combination of intellectual and statistical methods, as the CARMEN project and others do.

An interesting point in Chen et al. (1996) is the report that the term associations users made in cross-domain searching were context-driven: "...Based on our protocol analysis, we found that several contexts for these similarities and differences existed, including, two genes were identified by similar (or different) experimental strategies; their cellular structures had similar (or different) composition; ... genes manifested similar or different phenotypes; genes or proteins had similar or dissimilar sequences (homology) or contained similar motifs or domains; proteins or genes performed similar (or dissimilar) functions;... " This paper contributes to a better understanding of such phenomena. Even though semantic heterogeneity of terminological resources has frequently been referred as a problem, a systematic analysis of its intellectual basis and structure has not been carried out. Krause (2000) writes: "... the information market over the past twenty years... mainly views the development in distributed databases, user interfaces and the Internet as technological improvements and problems of standardization, without addressing the conceptual challenges involved.”

1.3 Thesaurus Mapping in One Domain

We regard thesaurus mapping as the process of identifying terms, concepts and hierarchical relationships that are approximately equivalent. It is a central process for merging thesauri, metathesaurus and cross-concordance construction, and thesaurus switching. This paper investigates the problems of finding appropriate equivalents, in particular focusing on issues related to polyhierarchies and the relationships between compound and non-compound terms.

![Figure 1. Scenario using correlated thesauri](image)

Figure 1. Scenario using correlated thesauri

The following assumes a general scenario (Figure 1) of a user addressing different digital collections using a particular thesaurus of choice, which is mapped to thesauri in other languages, to more specialized vocabularies, or to different versions of the thesaurus. We adopt the notion of a two-step process from Krause (2000a), which separates the vagueness introduced by thesaurus mapping (step one) from that introduced by the relationship between the user query and the document (step two). To separate these effects intellectually, each thesaurus is assumed to be consistent with the indexing of one or more collections, i.e. a correct user query to retrieve networked resources remains a fertile field for research and exploration.

Under these conditions, we attribute the remaining heterogeneity between different thesauri to (see Doerr 1996, p.3):

1. Different word use, due to different natural languages, the chosen language level, not semantically justified decisions in the selection of descriptors, and the degree of post- or pre-coordination of terminology. For example, the following terms all describe the same species: chaffinch (English), Buchfink (German), fringilla coelebs (scientific). Such differences are readily apparent when comparing terminology of thesauri in the same language with overlapping domains, such as the Art & Architecture Thesaurus from the Getty Foundation and the NDR Monument Type Thesaurus from English Heritage (formerly RCHME, in the following "NMR"), both in English.

2. Different coverage, due to different states of development, different scope and varying user needs. In particular, thesauri often develop some topics in far more detail than others. For example, something found under "dolls, Hopi" in
one thesaurus may be found under "Kachina" in another. Some place name thesauri may cover places only down to the country level, while another includes lower administrative areas and communities.

3. Different semantics, due to different conceptualizations. This occurs typically between thesauri from different languages, but it may also be due to different aspects of classification. For example, does "architecture" and its narrower terms denote types of buildings or the designing of buildings? Is "museum" a building or an organization?

4. Different semantic relations, often due to the enforcement of monohierarchies, but also due to different classification aspects. When polyhierarchies are not permitted, placement of terms that have two or more likely hierarchies (e.g., "chemical geology","decorative weapons") is based on considerations that are often not documented; two monohierarchical thesauri are likely to make different decisions about placement, which may not primarily be based on concept semantics.

We shall give several examples from the Art & Architecture Thesaurus (AAT) Getty AHP 1994, because we have been able to study it in detail and because it is an impressive example of a large, context-free thesaurus created with strict editorial principles ("context-free" means without a commitment to a specific application). Years after it began, it is now also a source to study what could be done better, a fact that does not diminish its extraordinary value (see also Soergel 1995). Other examples are taken from the intellectually created mapping of the Merimee "Thesaurus Architecture", to the AAT and the English Heritage Thesaurus (NMR) in the 1997 edition, as used in the AQUARELLE project and now being carried on in the HERIN Project.

Section 2 presents well-known ideas of concept-based thesauri in order to clarify several notions. First, we define the kind of mapping we mean to clarify the differences to other approaches. Second, we discuss distinct classes of "multilingual thesaurus", which is used in a fuzzy way in literature. Third, we refine the current notions of equivalence expressions, as given in ISO5964, in order to conform to certain logical requirements. Based on that background, we make a novel proposal for a methodology of mapping that allows for controlling vagueness in cross-thesaurus retrieval. This becomes increasingly relevant, as several projects are beginning to create such mappings on a large scale. Section 3, studies effects that may either impede the definition of equivalences between terms or between hierarchies, or impede the exploitation of semantic relations in the target thesaurus for query expansion. Chan (2000) regards thesauri as "a query expansion device", a virtue that should be preserved through mapping.

2 Application of Concept-based Mapping

Even though the ISO Guidelines for the establishment and development of multilingual thesauri (ISO5964) were published in 1985, it is only now that such mappings are being attempted on a larger scale. Since its publication, no specific methodology has been proposed about which terms should be correlated with equivalence relations and which not in the lack of exact equivalence, and how this would affect the query or information retrieval quality of the pair of correlated thesauri as a whole.

2.1 Concept-based mapping

In this section and some of the following we talk about sets in the mathematical sense. By objects we do not mean only documents. It may be anything in an electronically registered collection: a potsherd, a stool, a palace, an image, and a text. By sets we usually refer here to sets of such objects, typically defined by the sharing of one or more common properties. We cannot go into more details about Description Logic and similar theories here, and we can mention only the basic idea:

Under certain assumptions, preferred terms, so-called "descriptors", can be identified with concepts. Each concept in turn can be identified with the intention of a set of objects. In the sequence, we can transform the mapping problems into a mathematical problem about sets, i.e. terms are identified with the sets of objects they correctly classify (see Doerr and Fundulaki 1998 for details). "Correctly" is a question of user convention, and we assume that users can in general positively decide which term is correctly applied and which not. This assumption provides an absolute measure to compare concepts to thesauri even between multiple languages. As long as objects in a large enough database are classified in a well-defined way with two thesauri in parallel, set-relations between the concepts of both thesauri can be approximated automatically (as by Ambe et al. 1996). Any inconsistencies can then be reduced to human errors. Such assumptions are well known and basic to Description Logic (DL), e.g. Baader et al. (1999), Benjamins (1992), DL Web site, and implicit in many thesauri describing physical objects. In practice, not all subset relations may have been expressible in a thesaurus and term interpretation can be context dependent in a complex way, as will be discussed later. To make a clear distinction from statistical or neural network methods, let's define "concept-based mapping". The principles are:

1. A term is mapped to the associated set of objects which it correctly classifies (like the "interpretation function" in Description Logic).
2. The associated set of a broader (narrower) term (BT/NT) of some term is a superset (subset) of the associated set of the latter. In terms of DL, the broader term "subsumes" the narrower.
3. Some kinds of related term (RT) relationships can be identified with roles in the sense of DL, in particular the whole-relation (BTP) and functional relations.
4. The mapping between two terms is defined through the set-relations of their associated sets.

In this definition we use the relation notations BT, NT, RT, and BTP of ISO 2788. This definition of mapping is stricter than the term "cross-concordances" used by Krause (2000). It must be stressed that concepts are interpreted by descriptor use and not by comprehension of the term itself. Reuse of a thesaurus in a different context may change interpretation and require a redesign of the hierarchical relations (BT/NT), a complication often overlooked (see section 3).

2.2 About Multilingual Thesauri

Often any kind of relations between terms from thesauri in different natural languages are referred to as translations. In our opinion, translation in the proper sense differs from the concept-based mapping and cross-concordances in significant ways. In the AQUARELLE project, Dachelet (1997) proposed distinctions between different kinds of multilingual thesaurus. To clarify the differences we define the following classes of "multilingual thesaurus":

1. Translated thesauri. A thesaurus, where each concept is optimally interpreted in words of another or multiple languages, to allow speakers of those languages to understand better and use the concepts of this thesaurus more effectively. Note that such translations are in general not established indexing terms of the target language.
2. Correlated thesauri. An aggregate of multiple thesauri consisting of established indexing terms (concepts) of the respective user communities, and a set of concept-based mappings between the concepts from the different thesauri of that aggregate. The mappings serve as replacements of the original terms in queries sent out against multiple databases. Each database uses one of the thesauri of such an aggregate (see e.g. the AQUARELLE project).
reduction is done in order to obtain equivalent results from all databases to the degree possible. Note that the mapped concepts are in general not good translations of each other, because they do not interpret the other concept, but the associated sets of the mapped concepts contain common objects. This definition can be relaxed to weaker kinds of correlations. It is consistent with ISO 5964, but obviously not restricted to the case of different natural languages.

3. **Interlingua.** A thesaurus made out of concepts that are created by fusing each cluster of similar concepts from different social groups into a new concept (by analogy with the solution in machine translation, see e.g. Hutchins (1995)). In this process, one term from each user group is attached to the new concept as the identifier to be used by this group. The interlingua provides the sharing of concepts between social groups, e.g. as a legal basis used by the European Commission like the EBTI. Note that the interlingua may not contain any of the original concepts of any user group; it contains a set of compromises to remove interpretational differences. Its concepts may again be translated and correlated to other thesauri.

Figure 2 illustrates symbolically the three types of multilingual thesauri. It alludes to the example of the correlations created by the French Ministry of Culture between the Merimee "Thesaurus Architecture" and the English Heritage thesaurus (NMR). For example, Merimee's tennis = NMR's tennis court AND tennis club; megalithe has a narrower equivalence standing stone. We have added hypothetical translations and a hypothetical interlingua. Sometimes a system of well-defined concepts from one group is adopted as a whole by a user group of another language. For example, the Library of Congress Subject Headings are used in our Greek university library and in many libraries of other smaller language groups. Similarly, there exist translations of the American Art & Architecture Thesaurus to Dutch and Spanish, and others are planned. In such cases the concepts in the translated thesaurus can become the interlingua, with the benefit that it contains at least one's original concepts. In this paper, we are interested primarily in correlated thesauri, which are also the aim of ISO 5964.

**Figure 2.** Demonstrating different notions of a multilingual thesaurus in one context

### 2.3 Equivalence Expressions

Equivalence expressions similar to those in ISO 5964 are used with increasing frequency for thesaurus mapping: in the Merimee "Thesaurus Architecture", CARMEN, MACS, the HEREIN Project and others. Based on the idea of concept-based mapping and on the argument that correlated thesauri should serve equivalent retrieval results across systems employing different terminological resources, we have proposed that the expressive power of the mapping should be at least equivalent to the expressive power of the search paradigm. Otherwise, the user could express better queries in each target system than the mapping mechanism could provide. Doerr and Fundulaki (1998) investigated the mapping equivalent to the Boolean expressions (AND, OR, NOT) foreseen by the Z39.50 protocol. We found that a slight extension to IS05964 is sufficient to achieve equivalence expressions with the expressive power of Boolean queries. For that purpose, we interpret the equivalence expressions of ISO5964 as concept-based mappings, i.e. as set relations of the associated sets of objects. This seems to be justified by the Venn-diagram-like illustrations in the ISO5964. We make the following interpretations and extensions:

- "partial equivalence" should become "broader equivalence" (is subset of) or "narrower equivalence" (is superset of).
- "exact equivalence" is interpreted as "same set as".
- "inexact equivalence" is interpreted as "overlaps with".
- "single to multiple equivalence" should become "equivalence" to "compound" where "compound" is a Boolean expression of target terms with AND, OR, NOT and "*" is either "exact" or "broader" or "narrower".

The Getty Information Institute (1996) proposed the symbolism of a broader ("<") and narrower (">") equivalence, AND combinations ("*") and OR combinations ("&"), as does the on-going HEREIN project. Other Boolean expressions have not yet been proposed. Boolean expressions are interpreted as intersections, unions and complements of the associated sets. The UMLS Metathesaurus, due to its tighter coupling, uses implicitly exact, broader and narrower equivalences, as well as explicit term combinations ("Associated Expressions") using AND combinations.

So far, these equivalence expressions provide a means to express initial query terms in terms of any target thesaurus. Obviously, any Boolean combination of terms in the initial query can be converted into a Boolean combination of target terms (see below). Figures 3 and 4 demonstrate the semantics of equivalence expressions with Boolean compounds. An example for Figure 3, the French term bergerie has the exact equivalence to "sheep barns OR sheep folds" in the AAT. The first common broader term of both terms in the AAT is single built works. The obvious broader term animal housing is the broader term in AAT only to sheep folds, probably because of its monohierarchy design (see section 3.3). For Figure 4, please see the list in the Appendix. In Figure 4, the dotted circle on the right-hand side indicates where the approximated concept would appear in the target hierarchy, under the assumption that the BT relation expresses subsumption.
2.4 Methodological Aspects

Above, the arguments are constrained to the controlled vocabulary situation: concept-based mappings, derived from ISO5964, used to create correlated thesauri. Under these restrictions the effect of the following methodological arguments can be evaluated theoretically. These restrictions are realistic. For example, the HEREIN project will connect databases about material cultural heritage with thesauri correlated in the style of ISO5964. The precision of manual object classification is precise in the way assumed in section 2.1. Other obvious cases of precise classification are the use of place name authorities (gazetteers) like the Thesaurus of Geographic Names (TGN) and cultural period authorities. Professional users in cultural heritage administration (as investigated in the AQUARELLE and Term-IT project) and many other disciplines require stricter standards for recall and precision than general users seeking information on the Internet. The optimization of the recall/precision ratio usual in information retrieval does not satisfy the needs of a statistical survey. Chan (2000) also refers to the requirement for recall and precision as distinct: “Subject access tools are used to enable optimal recall... to enable optimal precision...”. In the free-text situation, basically the same arguments presented below should hold, but the effect will not be so explicit because several factors introduce additional vagueness. The question of when the effect of more elaborate correlations vanishes in the vagueness coming from other sources is interesting, but beyond the scope of this paper. The same holds for the question of whether the effort to create mappings intellectually or semi-automatically is affordable or not. We are satisfied here with the fact that people are increasingly undertaking that effort.

To illustrate the relevance of the following, Table 1 presents some statistics about the carefully produced equivalence expressions from the 1997 editions of the French Merimee “Thesaurus Architecture” to the AAT and NMR thesaurus. Table 2 compares the frequency of equivalence expressions among these three thesauri.

Table 1. Characterizing three thesauri

<table>
<thead>
<tr>
<th>Thesaurus</th>
<th>Domain</th>
<th>Language</th>
<th>Hierarchies</th>
<th>Pre-coordination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merimee</td>
<td>Western architecture</td>
<td>French</td>
<td>mono</td>
<td>high</td>
</tr>
<tr>
<td>ATT</td>
<td>Art and Western architecture</td>
<td>English</td>
<td>mono</td>
<td>low</td>
</tr>
<tr>
<td>NMR</td>
<td>Western architecture</td>
<td>English</td>
<td>poly</td>
<td>high</td>
</tr>
</tbody>
</table>

Table 2. Distribution of equivalence relation types in the Merimee mappings

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Total number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merimee</td>
<td>1336</td>
<td>100%</td>
</tr>
<tr>
<td>Number of terms with equivalence to AAT:</td>
<td>795</td>
<td>59%</td>
</tr>
<tr>
<td>Number of exact equivalence to AAT:</td>
<td>687</td>
<td>59%</td>
</tr>
<tr>
<td>Number of partial equivalence to AAT:</td>
<td>119</td>
<td>15%</td>
</tr>
<tr>
<td>Number of OR combinations to AAT:</td>
<td>26</td>
<td>3%</td>
</tr>
<tr>
<td>Number of AND combinations to AAT:</td>
<td>196</td>
<td>25%</td>
</tr>
<tr>
<td>Number of terms with equivalence to NMR:</td>
<td>735</td>
<td>55%</td>
</tr>
<tr>
<td>Number of terms with equiv. to AAT and NMR:</td>
<td>634</td>
<td>48%</td>
</tr>
<tr>
<td>Number of exact equivalence to NMR:</td>
<td>596</td>
<td>78%</td>
</tr>
<tr>
<td>Number of partial equivalence to NMR:</td>
<td>165</td>
<td>22%</td>
</tr>
<tr>
<td>Number of OR combinations to NMR:</td>
<td>86</td>
<td>11%</td>
</tr>
<tr>
<td>Number of AND combinations to NMR:</td>
<td>8</td>
<td>1%</td>
</tr>
</tbody>
</table>

In the mapping to the AAT, AND combinations mainly reflect post-coordination rules. Of the 199 so-called AND combinations used by Merimee to map to the AAT, at least 174 turned out to be role restrictions rather than true AND combinations (see...
section 3.3). They are listed in the Appendix. NMR is far more detailed and pre-coordinated, therefore OR combinations dominate, with up to 61 terms combined. The AND combinations to NMR follow the logic of Figure 4. As one team has created the above equivalences, the differences must be attributed to the nature of the target vocabularies and not to differences in the practice of the editors.

Some 40% of the Merimee terms are not mapped, with no indication at all which terms they relate to in the other thesauri. For example, the French term EDIFICE FUNERAIRE has no equivalence to the NMR. Its narrower terms MAUSOLEE and OSSUIRE have one equivalence each: mausoleum, ossuary. Obviously, the broader term is mausoleum in NMR: funerary site could be a broader equivalence of EDIFICE FUNERAIRE, but that has not been declared. Probably the editors felt that it would be "too far". This is an example of the proposal illustrated in Figure 5. Table 2 illustrates the complexity of thesaurus mapping, and shows that mappings not created with a well-founded methodology for networked information retrieval do not provide the necessary qualities, as will be discussed in the rest of this paper. Moreover, we think that providing such quality would not require much greater effort than that invested in a mapping like that presented above.

Figure 5. Demonstrating semantics of term inclusion

Now consider the question of whether an optimal mapping is possible. If someone uses an arbitrary set of equivalence expressions for correlation, the existence of an equivalence link for some query terms is not guaranteed. So a query using such a term would simply fail against the respective target. Even if an equivalence expression exists, the relation between the intended and the actual query with replaced terms is a priori unpredictable. Equivalence relation creation in a larger environment (Hutchins 1995) causes some "combinatorial explosion", therefore "switching languages" are proposed as intermediaries. The results of such subsequent mappings are even more undefined. We therefore propose to approximate concepts of a source thesaurus systematically by confining them within the nearest broader and narrower concepts of the target thesaurus (Figure 5). The idea can be stated in the following rules:

1. If a concept c of thesaurus A has no exact equivalence to any concept of thesaurus B then: for this concept c at least one broader equivalence and at least one narrower equivalence to some appropriate concepts in thesaurus B should be declared (see Figure 5).
2. The above broader equivalence should be minimal, i.e. there should be no other term combination in B which is broader than concept c and narrower than the given broader equivalence.
3. The above narrower equivalence should be maximal, i.e. there should be no other term combination in B which is narrower than concept c and broader than the given narrower equivalence.
4. The user has the chance to specify if the transformed query is going to preserve recall or precision.

If Rule 1 holds, we consider this a "complete mapping." If Rules 1, 2, and 3 hold, the correlation is considered to be an "optimal mapping." Obviously, choosing a broader equivalence for a single term in a Z39.50 protocol access profile element will preserve recall, i.e. return a set containing the relevant items plus some non-relevant ones. Negation turns things over: the narrower equivalence will preserve recall, and the broader precision. In practice, both recall preservation and precision preservation are needed; e.g. recall for statistics (under the assumption, that uncorrelated items can be sorted out in a second step) and precision for rapid discovery of relevant items.

Rule 1 defines a notion of "completeness" of the mapping, that is, if Rule 1 holds, the replacement of any query term is possible. of course, it may be impossible to find any narrower equivalence, in which case we use the empty concept ("bottom" in DL). In this case, the query returns no result, which is consistent. In the case of negation, however, it would return the universal concept ("top" in DL), that is, the whole target base. This is a case to be prohibited. Even though, such a situation can be tolerable in a query refinement cycle, as the actual intermediate results may not be transferred. In rare cases it may even be impossible to find a broader concept, in which case one must map to the universal concept. This situation could be avoided if thesaurus providers agree to share some high-level concepts. In addition, other AND combinations as they appear in a typical user query may "absorb" the universal concepts and return reasonable results. If the result of a translated query would be the whole target base, the user should be informed and given the choice to cancel the query. We see here an area for pragmatic solutions in the framework of application development. Note also that concept inclusions propagate without problem through multiple intermediate translations.

The above qualitative reasoning holds in this form only for simple conjunctive queries. In general, the problem can be reduced to a query containment problem, if the database schema plus terminology are interpreted as a schema altogether. For the latter, elaborate theories about the complexity and decidability of query containment exist (Calvanese et al. 1998 on PODS), which have to be applied on a case-by-case basis. Calvanese et al. (1998) on KR ’98 present a fairly general, unified framework for information integration from heterogeneous sources.

Rules 2 and 3 above define a notion of an "optimal mapping" in the sense that no closer mapping can be found with these kinds of expressions. If one actually chooses Boolean combinations rather than only the primitive concepts for the term inclusion, things may become algorithmically complicated and can go beyond the capacity of normal expert insight. Not even typical Description Logic implementations as CLASSIC or Fact answer such questions. On the other hand, if the mapping is close to optimal, the vagueness control still exists. Here is an area for further applied research.

Finally, the use of Boolean compounds poses some more methodological questions. Whereas other equivalence expressions can be read (anti-) symmetrically (reciprocally), e.g. narrower equivalence reads reciprocally as broader equivalence, a Boolean compound can not be easily interpreted in the opposite direction (see Figures 3 and 4). Thesaurus or DL tools don’t even normally indicate which concepts are used in a Boolean compound. Our laboratory has implemented a research prototype (Ntoas 1999) on top of the thesaurus management system SIS-TMS (Doerr and Fundulaki 1998a), which indexes expressions
containing Boolean operators and role restriction ("restrict [p,C]") and relates them to their immediate broader and narrower terms. Future research could address the question of the degree to which Boolean compounds or more complex DL expressions of some mapping can be exploited to calculate equivalence expressions in the opposite direction.

Summarizing, under the given assumption the proposed mapping methodology allows for the propagation of any query to collections classified with thesauri that have been mapped to one another. The query will not fail for any given query term, except if the whole database should be returned. As concepts do not precisely match, the results cannot always be equivalent. Rather, following the choice of the user, a (smallest possible) larger result or a (largest possible) smaller result can be returned, which could be further refined through post-processing steps. Such results can be used for statistical purposes.

For the free text situation (see section 1), inexact equivalence as defined in ISO5964 should be quite useful in balancing recall and precision. I guess, however, that the appropriate generalization of an inexact equivalence for full-text retrieval is a correlation based on associated relevance weights. We shall not follow this subject further here.

All of the above is based on the ideal assumption that the correlated thesauri follow the same rules and that their BT relations are complete, consistent, and follow the same logic. If this is not the case, some consistency checks can be carried out. At least it can be verified if manually or automatically derived equivalence relations won’t cause cycles with the given BT relations on either side, i.e. if some concept seems to include one of its broader concepts. In some cases the search for equivalences may reveal missing additional BT relations on either side, as sheep barns BT animal housing above. The rest of this paper is devoted to thoughts about the reasons for inconsistencies between hierarchies, the problems appearing in reality.

3 Heterogeneities of the Hierarchical Structure

If two correlated thesauri use subsumption hierarchies (or IsA relationship) and declare explicitly all direct subsumption relations between their primitive (non-compound) concepts, many nice applications can be done. The transitivity of subsumption allows expansion of query terms into their narrower terms to arbitrary depth, in particular by correlated concepts of another thesaurus and their narrower terms. This allows switching use from one thesaurus to another, e.g. to a more specialized one. Thus a general high-level thesaurus can be federated with a series of application-specific thesauri. Further, the subsumption relations between all terms of two thesauri can be calculated from a complete mapping in the above sense, and eventual logical inconsistencies can be reduced through human errors and eliminated. In practice, however, term hierarchies often (1) do not express subsumption, (2) are ambiguous, or (3) do not express all immediate subsumption relations. Some reasons and possible solutions are analysed below.

3.1 Hierarchical relation without subsumption

Traditionally, thesauri were printed books, and the hierarchies were used as an association mechanism to lead users most effectively to a concept for which he or she does not know the term. The sequencing into book pages does not foster the use of polyhierarchies. Hence, thesaurus hierarchies were like decision trees than semantic relations. With computers, representation restrictions are less limiting (Soergel and Jenkins 1999). Fascinating in this context are Ranganathan’s (1965) classical considerations about the obstacles the "notational plane" causes to the development of the "ideal plane". The traditions from editing printed books are not easily overcome, however. So often any hierarchical relation is messed up with subsumption, as they are equally useful for user guidance. In our opinion, user guidance and semantic relations are not all the same and should coexist in the same thesaurus.

ISO 2788 still regards the part-whole relation (BTP) as a kind of Broader Term relation, whereas e.g. the AAT Editorial Manual (1998) already regards them as a kind of Related Term (RT, "Code 2B"), and no longer as subsumption. Another example is the inclusion of geographical areas in place name thesauri (e.g. the Thesaurus of Geographical Names (TGN)). Obviously, France isA Europe does not hold. Different hierarchical relations hold for temporal intervals and cultural periods, even though there are still few examples of thesauri about periods. The CIDOC CRM ontology (Doerr and Crofts 1999) refers to these four relations as forms part of. These four relations can be explained by their extensions to different related sets, i.e. sets of points on the surface of earth, in an object, on the time-line, and in space-time. Hence they inherit the partial order relation from the subsumption of the respective sets, form (poly)hierarchies, and are therefore frequently mistaken for broader terms. Transitivity does not extend among them, however (see Motschnik 1993) about limited transitivity between different part-whole relations, and therefore they cannot be mingled. Nevertheless, expansion of query terms, e.g. from object types into their parts, can be useful if explicitly required.

Another source of confusion are the semantic relations within a set of derived concepts, parallel hierarchies as described by Soergel (1995) or the DL role restriction (Baader et al. 1992). For example, even though "Greece IsA Europe" does not hold, "Greek person IsA European person" does hold. In this case, Greek person is interpreted as Person.who lives in: Greece. Who lives in can be seen as a DL role, here restricted to Greece. Another example: even though "bridge construction IsA bridge" does not hold, "book about bridge construction IsA book about bridges" may be regarded as valid. The use of terms in a specific database field may hide a concept derivation, e.g. when object names are used as subjects or place names as nationalities. Editors may introduce in their thesaurus the subsumption relations correct for that use, i.e. of the derived concepts. Out of context those can be completely wrong.

In particular, subject headings often refer to physical objects (e.g. objects in museum collections), but their hierarchies cannot be directly used for object classification, causing frequent disputes between librarians and museum curators. Characteristically, the Getty Information Institute has rearranged large amounts of terms from the LCSH (Library of Congress Subject Headings) into the AAT, a thesaurus mainly about physical objects (its Object facet, except for the hierarchy information objects). We regard it as worthwhile to investigate under which conditions associations such as the above can cause subsumption relations in their derived concepts. Related to this is Wetty and Jenkins’ (1999) thorough study on modelling subjects. We would not expect the opposite, i.e. that concept derivation from a subsumption hierarchy may not preserve this hierarchy in the derived concepts. At least theoretically it should not happen (Nees 1999). Subject terms used for library cataloging are sometimes interpreted as applying to books which cover the breadth of the term. For example, biology of mammals would be used to denote books about the biology of all mammals, rather than some mammals. In this interpretation, narrower terms are not subsumed.

Finally, thesauri such as the "dmoz" project, which lets users act freely, end up with associations motivated by any contextual link, and even may contain cycles and other relationships that break thesaurus rules. For example, both "Top: Arts: Classical Studies: Journals" and "Top: Arts: Classical Studies: Academic Departments" are declared as narrower terms of "Top: Arts: Classical Studies" (DMOZ).
Summarizing, there are hierarchical relationships used in thesauri that are not subsumptions. They have to be clearly marked as being of different nature so that correct reasoning can be done. Otherwise, screws may be taken for cars, villages for nations, Andorra for a continent, etc. The use of terms in a specific database field may hide a concept derivation, and out of context hierarchies made for that use could be wrong. Therefore, the assumed semantics of hierarchical relations should be made clear before thesauri are correlated and should be communicated as thesaurus metadata or clear notations for the relationships.

3.2 Context-induced ambiguity

Terms and concepts often reveal a polysemy which is disambiguated by the context in which they are used. English, for instance, is full of so-called homonyms or contrastive ambiguity (Pustejovsky 1995), like: orange (color) and orange (fruit); pink (color) and pink (vessel); column (architectural element) and column (text arrangement). Even though some older thesauri maintain hierarchies on word-based hierarchies, concept-based hierarchy organization prevails in modern computer science. Consequently, the concepts have to be disambiguated. For example, in the AAT, a domain determinator like color in orange (color) disambiguates the concept. The actual word (e.g. orange) can be attached as a non-preferred term or synonym to all possible meanings. WordNet (Miller et al. 1993), for example, uses many-to-many relations between words and concepts, the most consistent approach to represent the real relation.

So far, the problems of homonymy seem to be solved by this so-called sense enumeration (Pustejovsky 1995). Each sense of a term represents a concept independent of contextual influence, and subsumption hierarchies can be designed independent of use. Homonymy is a language-specific feature, i.e. the different senses of one term in one language are normally translated into different terms in another language.

There are, however, the more subtle cases, which Pustejovsky calls complementary polysemy, an expression of the dynamic power of the concept formation behind our languages. For example, is *door* an object or an opening? Is *neck* a part or a place on a body? Is school an organization or a building? These terms are typically translated one-to-one into other languages for the same set of meanings. Pustejovsky introduces the notion of *qualia*, the different aspects that may cause a word to change meaning in context. I have the impression, that this polysemy may be intrinsic to the concept itself. He talks about the *Qualia Structure of nominals*, which he analyses in the following main categories (referring also to Aristotle’s notion of modes of explanation):

- Constitutive: the relation between an object and its constituents (material, weight, parts and component elements)
- Formal: that which distinguishes the object within a larger domain (orientation, magnitude, shape, dimensionality, color, position)
- Telic: purpose and function of the object (purpose an agent has in performing an act; built-in function or aim which specifies certain activities)
- Agentive: factors involved in the origin or “bringing about” of an object (creator, artifact, natural kind, causal chain)

This analysis has a striking similarity with criteria for the term specialization we found from an empirical study (Doerr and Kalomoirakis 2000) in the guide terms of the AAT. Terms about man-made objects, for example, systematically seem to have functional (telic), morphological (formal), and constitutive aspects. We have the feeling that a few *qualia* dominate in each application. If this is the case, they can be formalized and their consequences identified and communicated as thesaurus metadata.

A clear distinction between concept and term, as in the case of contrastive ambiguity, cannot be made, and Pustejovsky regards sense enumeration as impractical. Three problems arise from that:

1. Often thesauri are not made with the full breadth of application in mind. Consequently, one or other aspect may be neglected. For example, the AAT regularly defines classifications by form and by function, but it seldom relates more specific concepts to both categories. Figure 6 demonstrates the necessity of multiple generalizations using the example of *foils*.
2. A thesaurus made to capture a specific aspect (e.g. in the AAT, the morphological aspect dominates) may provide insufficient or even wrong hierarchies if used under another aspect. Figure 6 symbolizes with different colors how each aspect may give rise to different kinds of narrower term relations. This subject may deserve further investigation.
3. The coverage of a concept may vary under these aspects. For example, the functional versus morphological aspects of a sword: is a children’s wooden toy sword a sword? There were quite functional wooden swords in Japan. Shops are full of *minoan sword imitations*. In the Archeological Museum of Heraklion, there is a presumed part of a non-functional Minoan sword, pretty sharp, which seems to have been an instrument for dangerous artistic exercises. And finally, there are sword-like letter openers.

![Diagram of multiple broader terms under multiple aspects](https://journals.tdl.org/jodi/index.php/jodi/rt/printerFriendly/31/32)

Figure 6. Multiple broader terms under multiple aspects

To the best of our knowledge, the problems arising from complementary polysemy for thesaurus design has not yet been studied. Under the above considerations, it is useful to make the *qualia* of the use of a thesaurus explicit and to add such characteristics to thesaurus metadata. For example, a *hammer* in the morphological sense, as archeologists would classify...
items, has nothing in common with a steam hammer. This may even provide incorrect broader terms for the functional aspect an engineer needs, who sees both concepts closely related as types of impact devices.

On the other side, thesaurus editors should systematically take into account other aspects of use and identify the additional broader term relations the other aspects require, as long as they are not contradictory. Maybe a better solution would be to make BT relations aspect-specific, as suggested by the colors in Figure 6. A similar situation is shown by Putievskiy (1995) on p. 145. Of course, we are aware that this may become quite labor intensive if no automatic methods are found. The whole topic leaves many questions open. Chen et al. (1996) on the context-driven associations cited in section 1 confirms our impression that these problems are highly relevant for thesaurus mapping. Such difficulties seem to be often ignored by technicians or taken as unavoidable vagueness of human argument.

3.3 Missing subsumption relations

Besides reasons of different contexts of use, we see the enforcement of monohierarchies and post-coordination rules as the major reasons for missing subsumption relations. As mentioned above, monohierarchies were preferred as long as the predominant medium for thesauri were books. A nice example is the colorant hierarchy of the AAT. Look at the position of crimson lake and carmine in this hierarchy in Figure 7.

- <materials by function>
  - <colorant
    - - - pigment
    - - - - pigment by color
    - - - - - red pigment
    - - - - - organic red pigment
    - - - - - - crimson lake

- <materials by function>
  - - colorant
    - - - <colorant for dye and pigment
      - - carmine

Figure 7. Position in hierarchy of the AAT terms "crimson lake" and "carmine"

There are three inherent aspects: functional form (pigment, dye, lake...), appearance (red, blue, brown...), production or provenance (artificial, organic). The above sequence seems to have placed the two terms arbitrarily. One could as well start with color or provenance. Carmine does not appear at all under its characteristic color, because priority was given to other prominent features (carmine is used for dye and pigment). The scope note of crimson lake states: "Deep, transparent, ruby-red lake pigment with bluish undertone, made from kermes, a natural dyestuff of insect origin; carmine, a better pigment introduced in the 16th century, became its chief competitor. MAIER." Hence carmine is a red organic pigment, but does not appear under that term and is far away in hierarchy from a very similar one.

Imagine a thesaurus federation: a thesaurus may have a leaf node of organic colorant and we would like to employ the AAT as the source for more specific terminology. Even though all concepts are in principle in the AAT, and there is no difference in conceptualization, we cannot continue from organic colorant to narrower terms in the AAT. In a student project, we have created an experimental colorant hierarchy, where each colorant is put directly under three broader terms: functional, color, and provenance terms. The complete result is difficult to show graphically, but browsing is effective, because on descending one branch or the other one comes down to the correct end and no possible broader term is missing.

This brings us to the other point: post-coordination. Obviously it is quite inefficient to define all combinations of terms like artificial inorganic red pigment, synthetic organic green pigment, etc. Therefore, thesauri like the AAT and the German subject headings Schlagwortsnormdatei (SWD) use rules to combine terms dynamically. For reasons of simplicity, only a "+" and "&" signs are used, both heavily overloaded with different interpretations. For example, "factories + grinding" in the AAT means a "factory which does grinding"; i.e. nothing more than a mill. The term mill has been sacrificed to post-coordination (it is a decoordinates subject in the AAT terminology), as are many other useful terms, as can be seen in the Appendix from the Merimee mappings. See also Soergel’s (1996) extensive analysis of such problems in the AAT. In current practice post-coordination suffers from three problems:

1. Unclear semantics and no consistency control.
2. Under the current convention, post-coordinated terms are leaf-concepts, because no mechanism is foreseen to relate another established term to a post-coordinated compound, as a narrower term or other.
3. No indication of which parts of hierarchies it makes sense to combine.

Hence it is fairly difficult to reconstruct the missing terms and their broader terms, a major obstacle to thesaurus mapping and federation - but neither can precoordination be seen as a solution, because it overloads thesauri. Problems 1 and 2 are appropriately solved by DL. Languages like GRAIL (see also Rector et al. 1997) provide a user-friendly syntax. Problem 3, however, is an open issue. Rich DL implementations do not, for obvious reasons, allow browsing through all possible concept combinations.

On the assumption that it makes sense for specific parts of hierarchies to be post-coordinated with specific relative roles, Ntoas (1999) designed a mechanism in which the user can declare that a specific concept or subhierarchy can be refined by restriction of a specific role to another subhierarchy. For example, factories which do grinding, etc. In the sequence, the user can browse through the virtual hierarchy induced by the subsumption properties of the respective processes. Further, explicitly declared natural concepts like mill will appear at their natural position in the virtual hierarchy, and narrower terms of mill can be added, which is impossible in the AAT. Boolean combinations, which do not define or lead to a natural concept, were not included in the browsing. They play a minor role in natural concept formation and are easily handled and understood by users.

Roles were taken from concepts in the CIDOC CRM (Doerr and Crofts 1999) ontology and from roles we found to be implicit in AAT terms. The semantic relations of the UMLS Semantic Network are another source of relevant roles, not only for the medical domain. It seems that a few relatively generic roles may actually be sufficient for most cases. The mechanism was verified with term combination from the equivalence expressions between the 1997 editions of the French Merimee "Thesaurus Architecture" and the AAT listed in the Appendix. For example, the compound "factory & owner’s & houses" is interpreted as...
houses which has owner: Person, which is owner of: factory”; wood & roofs as "roofs which consists of: wood”; umbrellas & factories as "factories which produce: umbrellas”, etc.

The above examples and Figure 6 demonstrate that post-coordination is useful at all levels of hierarchy, but that there must be a mechanism to embed natural concepts in post-coordinated schemes. We regard mechanisms simulating hierarchies of post-coordinated concepts as necessary to mediate effectively between pre- and post-coordination in thesaurus mapping and federation.

3.4 Summary

Section 2 showed that a suitable methodology to create thesaurus mappings can provide well-defined global recall and precision qualities for transitions between thesauri, that have so far not explicitly been considered. Thus we made assumptions that are realistic but often not present. In this section, we have studied effects that may undermine those assumptions. Some can be avoided, either by better awareness of the thesaurus providers or by specific reasoning services. In other cases, information about implicit assumptions can help avoid comparing incomparable structures. Therefore, we have repeatedly proposed that certain thesaurus characteristics be documented in metadata, to avoid semantic heterogeneity conflicts and to facilitate interoperability of reasoning mechanisms. These characteristics are:

1. The use and notation of different kinds of hierarchical relations, which are not transitive among each other.
2. The use of subsumption relations, not for the concept itself but for use-dependent derivatives.
3. The aspect (qua) under which subsumption hierarchies are created. This point needs further investigation.
4. The use of monohierarchies, provision of complete polyhierarchies, incomplete polyhierarchies
5. Details of post-coordination rules, which hierarchies can be combined, and the roles in use.

The Networked Knowledge Organization Systems (NKOS) group is an informal user organization devoted to the discussion of the functional and data model for enabling the interoperability of knowledge organization systems, such as classification systems, thesauri, gazetteers, and ontologies. For that purpose, is has defined a metadata format for KOS, the NKOS Registry. It provides a virtually complete description of the technical and administrative characteristics of a KOS, except that data about cardinality constraints are not required for all types of relationships. Aspects of use as above are not analysed in detail.

We propose to extend this metadata format by a suitable formulation of the five criteria above. The detail of semantic analysis we postulate here may appear to be too labor-intensive to pay off. This is, however, not a concern of a qualitative study like this. If one regards, for example, the incredible progress traditional dictionary writing has made by semi-automatic methods, I am quite confident that the near future will provide reasonable solutions. Already a scientific community has begun to concentrate on the issue of Ontology Learning (see conclusions of the OntoWeb Workshop). Therefore, we regard a good understanding of the intellectual problems of thesaurus integration as quite beneficial.

4 Conclusions

Many things can be done to bring forward information integration with thesauri. There is still a large gap between practitioners and scholars on one side, and theoreticians in knowledge representation and system engineers on the other. Whereas the practitioners administer the domain knowledge, the others have the technology to improve its handling. It is not easy for the practitioners to understand and appreciate the potential and limitations of technology, and often the theoreticians do not show particular understanding for intellectual problems in practice that do not directly conform with their models. We see a need for increased interdisciplinary empirical studies and verification of theoretical results, both to demonstrate the utility of theory to practitioners and to identify their limits and need for better theoretical understanding.

The results of several excellent implementations of thesaurus federations seem to have remained relatively unevaluated in terms of the real quality of concept mediation achieved. Some technology providers seem to see their task end at the point of installation and optimize their systems to work with any thesaurus. As we have tried to make clear in sections 2 and 3, we believe that thesaurus creators (scholars, experts and practitioners) have a responsibility to improve their methodology to meet the challenges of advanced technology (e.g. the completeness of hierarchies) and technologists have a responsibility to understand the complexity of the problem (e.g. contrastive polysemy). If this were to happen in a coordinated way, we could soon achieve a new quality of applications.

We believe that some concrete steps could be undertaken:

1. We see theory and practice that has advanced enough to implement query transformation services that preserve recall and precision in a controlled manner. To that end, a certain methodology of thesaurus creation and protocols for connecting to knowledge resources on the Internet should be agreed on. The interface between translation services and thesaurus mappings with clearly defined logical properties should be defined.
2. Methodological specifications for thesaurus providers should be developed that conform to respective technical requirements. Among these are: ISA-semantics of broader terms, completeness of broader terms, synonym creation for concept identification in free text, explicit aspects for hierarchy creation (functional, morphological, etc.)
3. Metadata for thesauri and the services employing mappings should sufficiently define the properties and methodological principles needed for coherent, dynamic information integration services.

We also see scope for applied research in the connection with formal ontologies (in the sense of knowledge representation models as semantic networks, KL-ONE-like data models or Description Logic) with thesauri. These are:

1. The identification of basic roles for concept formation (like the process-product-producer relation, using, made for, etc.). We believe this needs the cooperation of ontologists, linguists and empirical studies in terminological resources.
2. Methodological specifications for thesaurus providers should be developed that conform to respective technical requirements. Among these are: ISA-semantics of broader terms, completeness of broader terms, synonym creation for concept identification in free text, explicit aspects for hierarchy creation (functional, morphological, etc.)
3. Metadata for thesauri and the services employing mappings should sufficiently define the properties and methodological principles needed for coherent, dynamic information integration services.

Acknowledgements
I wish to express my thanks to the reviewers of this paper, and particularly to Linda Hill, for their valuable comments, which greatly helped to improve its quality.

References


Borgida, A. (1995) "Description logics in data management". IEEE Trans. on Knowledge and Data Engineering, 7(5):671--682


EBTI: A short description of the EBTI (European Binding Tariff Information) Thesaurus can be found in: http://www.bjl.be/2_3_1.htm


The HEREIN Project http://www.european-heritage.net/fr/Thesaurus/Contenu.html


https://journals.tdl.org/jodi/index.php/jodi/rt/printerFriendly/31/32
Table 3. All "&" combinations from the Merimee thesaurus to the AAT in 1997

<table>
<thead>
<tr>
<th>Count</th>
<th>French term</th>
<th>Relation</th>
<th>American term combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ACADEMIE</td>
<td>exact equivalence</td>
<td>academy &amp; buildings</td>
</tr>
<tr>
<td>2</td>
<td>AIRE DE CONCASSAGE</td>
<td>exact equivalence</td>
<td>crushing &amp; floors</td>
</tr>
<tr>
<td>3</td>
<td>AIRE DE LAVAGE</td>
<td>exact equivalence</td>
<td>washing &amp; floors</td>
</tr>
<tr>
<td>4</td>
<td>ARCHEVECHE</td>
<td>exact equivalence</td>
<td>bishop (prelate) &amp; palaces</td>
</tr>
<tr>
<td>5</td>
<td>ARCHEVECHE</td>
<td>partial equivalence</td>
<td>archbishop &amp; palaces</td>
</tr>
<tr>
<td>6</td>
<td>ARDOISIERE</td>
<td>exact equivalence</td>
<td>slate &amp; quarries</td>
</tr>
<tr>
<td>7</td>
<td>BASSIN</td>
<td>exact equivalence</td>
<td>artificial &amp; pools</td>
</tr>
<tr>
<td>8</td>
<td>BROSSERIE</td>
<td>exact equivalence</td>
<td>brush &amp; factories</td>
</tr>
<tr>
<td>No.</td>
<td>French Term</td>
<td>English Term</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------------------</td>
<td>-------------------------------------</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>BUREAU</td>
<td>exact equivalence factory &amp; offices</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>CABLERIE</td>
<td>exact equivalence electric cable &amp; factories</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CALVAIRE</td>
<td>exact equivalence calvary crosse &amp; monuments</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>CARTONNERIE</td>
<td>exact equivalence cardboard &amp; factories</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CHAIRE A PRECHER EXTERIERE</td>
<td>exact equivalence exterior &amp; pulpets</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>CHAMBRE DE COMMERCE</td>
<td>exact equivalence board of trade &amp; buildings</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>CHAMOISERIE</td>
<td>exact equivalence chamois &amp; factories</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>CHARPENTE EN BOIS</td>
<td>partial equivalence wood &amp; roofs</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>CHARPENTE METALLIQUE</td>
<td>partial equivalence metal &amp; roofs</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>CHARRETERIE</td>
<td>exact equivalence cart &amp; sheds</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>COLLATERAL</td>
<td>exact equivalence side &amp; aisles</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>CONCIERGERIE</td>
<td>exact equivalence porter's &amp; lodges</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>CONSERVATOIRE</td>
<td>exact equivalence drama &amp; schools</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>COURSIERE</td>
<td>exact equivalence wall &amp; passages</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>COUTELLERIE</td>
<td>exact equivalence cutlery &amp; factories</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>CRISTALLERIE</td>
<td>exact equivalence crystal (leadglass) &amp; factories</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>CROIX DE Cimetiere</td>
<td>exact equivalence cemetery &amp; crosses</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>DEPENDANCE</td>
<td>exact equivalence agriculture &amp; outbuildings</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>ECOLE D'AGRICULTURE</td>
<td>exact equivalence agricultural &amp; schools</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>ECOLE D'ART</td>
<td>exact equivalence art &amp; schools</td>
<td></td>
</tr>
<tr>
<td>29, 30</td>
<td>ECOLE DE DANSE</td>
<td>exact equivalence ballet &amp; schools OR dance &amp; studios (work spaces)</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>?dicule religieux chr?tien&gt;</td>
<td>exact equivalence christian &amp;</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>?dicile du g?nie civil&gt;</td>
<td>exact equivalence civil engineering &amp; buildings</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>EDIFICE RELIGIEUX CHRETIEN</td>
<td>exact equivalence christian &amp; religious buildings</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>ELEVATION INTERIEURE</td>
<td>exact equivalence interior &amp; elevations (building divisions)</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>ENCLOS FUNERARE</td>
<td>partial equivalence churchyard &amp; walls</td>
<td></td>
</tr>
<tr>
<td>36, 37</td>
<td>ENSEMBLE CASTRAL</td>
<td>exact equivalence castle &amp; complexes OR chateau &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>ENSEMBLE D'INDUSTRIE ALIMENTAIRE</td>
<td>exact equivalence food processing plant &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>ENSEMBLE D'INDUSTRIE CERAMIQUE</td>
<td>exact equivalence ceramic &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>ENSEMBLE D'INDUSTRIE CHIMIQUE</td>
<td>exact equivalence chemical &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>ENSEMBLE D'INDUSTRIE DU BOIS</td>
<td>exact equivalence woodworking &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>ENSEMBLE D'INDUSTRIE DU PAPIER</td>
<td>exact equivalence papermill &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>ENSEMBLE D'INDUSTRIE VERRIERE</td>
<td>exact equivalence glass &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>ENSEMBLE DE CONSTRUCTION</td>
<td>exact equivalence aircraft &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>ENSEMBLE DE CONSTRUCTION AUTOBILE</td>
<td>exact equivalence motor vehicle &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>ENSEMBLE DE CONSTRUCTION MECANIQUE</td>
<td>exact equivalence assembly plant &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>47, 48</td>
<td>ENSEMBLE DE CONSTRUCTION NAVALE</td>
<td>exact equivalence shipyard &amp; complexes OR naval shipyard &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>ENSEMBLE DE FABRICATION DE MATERIAUX DE CONSTRUCTION</td>
<td>exact equivalence building material &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>ENSEMBLE DE FABRICATION DES METAUX</td>
<td>partial equivalence &lt;metalworking plant&gt;&amp; complexes</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>ENSEMBLE DE PETITE METALLURGIE</td>
<td>partial equivalence machine shop &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>ENSEMBLE DU GENIE CIVIL</td>
<td>exact equivalence civil engineering &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>ENSEMBLE FORTIFICATION</td>
<td>exact equivalence fortification &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>ENSEMBLE FUNERARE</td>
<td>exact equivalence funerary buildings &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>ENSEMBLE METALLURGIQUE</td>
<td>partial equivalence &lt;metalworking plant&gt;&amp; complexes</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>ENSEMBLE TEXTILE</td>
<td>exact equivalence textile mill &amp; complexes</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>ESCALIER INDEPENDANT</td>
<td>exact equivalence freestanding &amp; stairs</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>ÉTABLISSEMENT CONVENTUEL</td>
<td>partial equivalence christian &amp; religious communities</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>ÉTABLISSEMENT DE BAINS</td>
<td>exact equivalence public baths &amp; baths</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>ÉTABLISSEMENT NAUTIQUE</td>
<td>exact equivalence boating &amp; clubhouses</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>ÉTABLISSEMENT PORTUAIRE</td>
<td>exact equivalence harbor &amp; buildings</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>FAIENCERIE</td>
<td>exact equivalence faience &amp; factories</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>FECULERIE</td>
<td>exact equivalence starch &amp; factories</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>&lt;fondations et sols&gt;</td>
<td>exact equivalence foundations (structural elements) &amp; pavements (surface elements)</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>FOUR A CHANVRE</td>
<td>exact equivalence hemp &amp; ovens</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>Fournil</td>
<td>exact equivalence bake oven &amp; buildings</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>GANTERIE</td>
<td>exact equivalence glove &amp; factories</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>GARAGE</td>
<td>exact equivalence automobile &amp; repairshop</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>GAZOMETRE</td>
<td>exact equivalence natural gas &amp; storage tanks</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>GLACERIE</td>
<td>exact equivalence mirror &amp; glass &amp; factories</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>HUILERIE</td>
<td>exact equivalence vegetable oil &amp; animal oil &amp; factories</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>LAC DE JARDIN</td>
<td>exact equivalence garden &amp; lakes</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>LAVABO DE CLOITRE</td>
<td>exact equivalence cloister &amp; lavaboies</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>LOCAL SYNDICAL</td>
<td>partial equivalence trade union &amp; buildings</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>LOGEMENT DE CONTREMAITRE</td>
<td>exact equivalence foremen's &amp; houses</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>LOGEMENT PATRONAL</td>
<td>exact equivalence factory &amp; owner's &amp; houses</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>LOGIS ABBATIAL</td>
<td>exact equivalence abbots' &amp; houses</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>MAISON AUX DIMES</td>
<td>exact equivalence tithing &amp; offices</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>MAISON MINIATURE</td>
<td>exact equivalence miniature &amp; houses</td>
<td></td>
</tr>
</tbody>
</table>

https://journals.tdl.org/jodi/index.php/jodi/rt/printerFriendly/31/32
80 MILLIAIRE exact equivalence Roman & milestones
81 MONTJOIE partial equivalence pilgrimage & markers (monuments)
82 OBSERVATOIRE exact equivalence astronomical & observatories
83 OUVRAGE D'ART partial equivalence civil engineering & structures (single built works)
84 PARFUMERIE exact equivalence perfume & factories
85 PASSAGE COUVERT exact equivalence carriage & porches
86 PASSAGE D’ENTREE exact equivalence carriage & passages
87 PERCEPTION partial equivalence tax collectors’ & offices
88 PLATIERE exact equivalence plaster & factories
89 PUISS D'AERAGE exact equivalence ventilation & shafts (spaces)
90 RAFFINERIE DE PETROLE exact equivalence petroleum & refineries
91 RAFFINERIE DE SUCRE exact equivalence sugar & refineries
92 ROBINETTERIE exact equivalence plumbing hardware & factories
93 SALLE DU THEATRE exact equivalence theater & auditoriums
94 SAVONNERIE exact equivalence soap (organic material) & factories
95 SECHOIR A CHATAIGNES exact equivalence chestnut & drying sheds
96 SECHOIR A MAIS exact equivalence corn & drying sheds
97 SUICRERIE exact equivalence sugar & factories
98 TEMPLE partial equivalence protestant & churches
99 TEMPLE PAIEN partial equivalence ancient & temples
100 TENNIS exact equivalence tennis & courts (buildings)
101 TONNELERIE exact equivalence barrel (container) & factories
102 TREFILERIE exact equivalence wire & factories
103 USINE A GLACE exact equivalence ice & factories
104 USINE D'ACIDE SULFURIQUE exact equivalence sulfuric acid & factories
105 USINE D'ARMES exact equivalence ammunition & factories
106 USINE D'ARMES partial equivalence weapon & factories
107 USINE D'ARTICLES EN MATERIE PLASTIQUE exact equivalence plastic & hardware & factories
108 USINE D'EBENISTERIE exact equivalence cabinetmaking & factories
109 USINE D'ELEMENTS EN MATERIE PLASTIQUE POUR LE BATIMENT exact equivalence plastic & building material & factories
110 USINE D'ELEMENTS PREFABRIQUES partial equivalence prefabricated & building material & factories
111 USINE D'EMBALLAGE ET CONDITIONNEMENT exact equivalence packaging material & factories
112 USINE D'EMBALLAGES EN MATERIE PLASTIQUE exact equivalence plastic & packaging material & factories
113 USINE D'EMBOITISSAGE exact equivalence stamping (forming) & factories
114 USINE D'ENCRES exact equivalence ink & factories
115 USINE D'ENGRAIS exact equivalence fertilizer & factories
116 USINE D'ESTAMPAGE exact equivalence cold & stamping (forming) & factories
117 USINE D'HORLOGERIE exact equivalence timepiece & factories
118 USINE D'IMPRESSION SUR ETOFFES exact equivalence cloth & printing & textile mills
119 USINE D'INSTRUMENTS DE MESURE exact equivalence measuring device & factories
120 USINE D'INSTRUMENTS DE MUSIQUE exact equivalence musical instrument & factories
121 USINE D'OUADE exact equivalence batting & factories
122 USINE D'OUVRAGES EN AMIANTE exact equivalence asbestos & factories
123 USINE DE BIMBELOTERIE partial equivalence wood & toy (recreational artifact) & factories
124 USINE DE BOISSELLERIE partial equivalence turning & factories
125 USINE DE BONNETERIE exact equivalence hosiery & factories
126 USINE DE BOUCHONS exact equivalence cork (bark) & factories
127 USINE DE BOUGIES exact equivalence candle & factories
128 USINE DE BOUTONS exact equivalence button (fastener) & factories
129 USINE DE BOUDAUDERIE exact equivalence gut & factories
130 USINE DE BRODERIE MECANIQUE exact equivalence embroidery & factories
131 USINE DE CAOUTCHOUC exact equivalence rubber & factories
132 USINE DE CELULOSE exact equivalence cellulose & factories
133 USINE DE CERAMIQUE exact equivalence ceramic & factories
134 USINE DE CHAPELLERIE exact equivalence hat & factories
135 USINE DE CHAUSURES exact equivalence shoe (footwear) & factories
136 USINE DE CHAUX exact equivalence lime & factories
137 USINE DE COLLES exact equivalence glue & factories
138 USINE DE CONSTRUCTION AERONAUTIQUE exact equivalence aircraft & assembly plants
139 USINE DE CONSTRUCTION AUTOMOBILE exact equivalence automobile & assembly plants
140 USINE DE CONSTRUCTION METALLIQUE exact equivalence metal & building material & factories
141 USINE DE CONTRE PLAQUE exact equivalence plywood & factories
142 USINE DE COSMETIQUES exact equivalence cosmetic & factories
143 USINE DE CYCLES exact equivalence bicycle & factories OR motorcycle & factories
144 USINE DE DECOLLETAGE partial equivalence screw & factories
145 USINE DE DENTELLE MECANIQUE exact equivalence lace & factories
146 USINE DE DETERGENTS exact equivalence detergent & factories
147 USINE DE FABRICATION DE MATERIAUX DE CONSTRUCTION exact equivalence building material & factories
148 USINE DE FABRICATION ET DISTILLATION DES GOUDRONS exact equivalence tar & refineries
149
<table>
<thead>
<tr>
<th>Page</th>
<th>Term</th>
<th>Exact Equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>USINE DE FERBLANTERIE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>151</td>
<td>USINE DE FEUTRE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>152</td>
<td>USINE DE FIBRE DE VERRE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>153</td>
<td>USINE DE FIBRES ARTIFICIELLES ET SYNTHETIQUES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>154</td>
<td>USINE DE FLACONNAGE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>155</td>
<td>USINE DE GRES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>156</td>
<td>USINE DE MATERIEL AGRICOLE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>157</td>
<td>USINE DE MATERIEL DE TELECOMMUNICATION</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>158</td>
<td>USINE DE MATERIEL ELECTRIQUE INDUSTRIEL</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>159</td>
<td>USINE DE MATERIEL FERROVIAIRE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>160</td>
<td>USINE DE MATERIEL FERROVIAIRE</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>161</td>
<td>USINE DE MATERIEL INFORMATIQUE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>162</td>
<td>USINE DE MATERIEL MEDICOCHIRURGICAL</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>163</td>
<td>USINE DE MATERIEL OPTIQUE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>164</td>
<td>USINE DE MATERIEL PHOTO</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>165</td>
<td>USINE DE MATERIES COLORANTES SYNTHETIQUES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>166</td>
<td>USINE DE MATERIES PLASTIQUES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>167</td>
<td>USINE DE MENUISERIE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>168</td>
<td>USINE DE MEUBLES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>169</td>
<td>USINE DE PAPIERS PEINTS</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>170</td>
<td>USINE DE PARAPLUIES ET CANNES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>171</td>
<td>USINE DE PARAPLUIES ET CANNES</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>172</td>
<td>USINE DE PASSEMENTERIE</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>173</td>
<td>USINE DE PEINTURES ET VERNIS</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>174</td>
<td>USINE DE PEINTURES ET VERNIS</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>175</td>
<td>USINE DE PORCELAINE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>176</td>
<td>USINE DE POTERIE</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>177</td>
<td>USINE DE PRODUIT TEXTILE NON TISSE</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>178</td>
<td>USINE DE PRODUITS CHIMIQUES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>179</td>
<td>USINE DE PRODUITS EXPLOSIFS</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>180</td>
<td>USINE DE PRODUITS PHARMACEUTIQUES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>181</td>
<td>USINE DE PRODUITS PHOTOGRAPHIQUES ET CINEMATOGRAPHIQUES</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>182</td>
<td>USINE DE QUINCAILLERIE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>183</td>
<td>USINE DE SERRURERIE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>184</td>
<td>USINE DE SOUFFRE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>185</td>
<td>USINE DE TABAC</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>186</td>
<td>USINE DE TAILLE DE MATERIAUX DE CONSTRUCTION</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>187</td>
<td>USINE DE TAILLE DE PIERRE POUR LA JOAILLERIE ET L'INDUSTRIE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>188</td>
<td>USINE DE TRAITEMENT DE SURFACE DES METAUX</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>189</td>
<td>USINE DE TRANSFORMATION DU LIEGE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>190</td>
<td>USINE DE VERRE CREUX</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>191</td>
<td>USINE DE VERRE PLAT</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>192</td>
<td>USINE DE VERRES OPTIQUES</td>
<td>partial equivalence</td>
</tr>
<tr>
<td>193</td>
<td>USINE LIEE AU TRAVAIL DU BOIS</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>194</td>
<td>VERRERIE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>195</td>
<td>VERRIERE EN COUVERTURE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>196</td>
<td>VESTAIRE D'USINE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>197</td>
<td>VILLA</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>198</td>
<td>VINAIGRERIE</td>
<td>exact equivalence</td>
</tr>
<tr>
<td>199</td>
<td>VOIRIE</td>
<td>exact equivalence</td>
</tr>
</tbody>
</table>