

Vida Yeung
Discussion 3
INFO 247
Summer 2021
8/1/2021

Discussion 3: Thesaurus Evaluation

I selected *The International Thesaurus of Refugee Terminology*, as archived in bits by the Bodleian Libraries of Oxford University. It is in English, French and Spanish. It has 28 facets, and the first five represented (along with class codes) are ‘Refugees and other groups of concern’ (1), ‘Population groups’ (2), and ‘Refugee protection’ (3). As with the remaining 25, each descriptor is capitalized.

(1) has two sub classes: ‘Refugees’ (1.01) and ‘Other groups of concern’ (1.02).

(2) has 18, the first three being: ‘Migrants’ (2.01), ‘Ethnic and minority groups- General’ (2.02), and ‘African ethnic groups’ (2.03).

(3) has 9, the first three: ‘Pre-flight and flight’ (3.01), ‘Entry, admission and reception’ (3.02), and ‘Refugee law’ (3.03).

Examining what I can find under the other main facets, I’ve noticed every descriptor, whether sub-facet or preferred term, is capitalized. Their choice of notation, facet number then period and two digits, tell me that the rest of the thesaurus also has only one layer of sub facets and no list could’ve exceeded 99 terms. Their ordering of main facets seem to be from what would most immediately concern someone evaluating a refugee-related situation to the least pressing details. Their sub facets seem to follow the ordering protocol. The hierarchy is presented in only one language at a time, with access to the others from the menu flush down the left side of the web page.

In the alphabetic listings individual terms' pages, each are presented in all three languages, with the accompanying relationships in only one at a time. However, a click on the term in one of the other languages should've hyperlinked users to the page with the relationships in that language. The French terms have the French national flag besides, while the Spanish one has a simple representation of the Spanish one. I can't find which flag would be for the terms in English because the Bodleian Libraries have not archived the alphabetically listed terms pages in the other languages. I'm certain, though, it would be the Great Britain one, because Oxford University is in England.

Information is listed in the order of TERM, Class Name (with notation from hierarchy), SCOPE NOTE, BROADER TERM- BT, NARROW TERM- NT, RELATED TERM- RT, and USE FOR- UF. These are how the indicators were presented on the page. Doing it this way indicates that the thesaurus's intended users were not only researchers or others knowledgeable about ongoing refugee crises around the world, but also students and curious/concerned citizens who are not well-versed in information science. Also, including the class name and notation from the hierarchy gives users another relational aspect of the term.

A social issue in constructing a multilingual thesaurus is one must consider who is in need/would be interested in the information. For example: a new thesaurus concerning chocolate would do well to include Vietnamese, to serve the country's emerging agricultural interest in the product, and in Chinese for the immense population of confectionary chef-hopefuls in China.

A cultural issue would concern syntax. For example: many languages have transnational dialects –like Arabic, French, and all the varieties of Malay- without unifying authority to set/enforce a standard –unlike Chinese and Japanese-; thus, when selecting terms the writer must

take into account which populations will use the thesaurus, and what terms/word combinations might be offensive or awkward to certain groups.

A most pressing political issue would concern access. For example: thesauri intended for a general Indian audience must include, alongside Hindi and English, the main languages of most if not all of the country's states and union territories to avoid alienating tens of millions of citizens.

Integrating cultures in multilingual thesauri borders on Herculean, because depending on the user population, some terms are simply related and not hierarchical; for example, in an English/Chinese thesaurus, 'rice' on the English side would be considered a NT of 'meal', whereas in the Chinese portion it would be an UF. Here is where a scope note (SN) would be useful on both sides, so even language learners from either culture can realize this difference. Some terms between cultures have no relation at all; for example, between a French/Japanese thesaurus, the latter might have 'head of state' USE 'emperor' NT 'deity', but the former –with a bloody history of annihilating their monarchy into the proverbial dust- wouldn't recognize any sovereign-righted individual as a national leader, let alone possibly identifying them as divine. A historical note (HN) can be used to clarify the differences between these cultures.

Writing multilingual thesauri before computers must've been like suffering a perpetual migraine. Poking through what's left of *The International Thesaurus of Refugee Terminology*, I couldn't understand how one could switch between the three languages on paper. Without hyperlinks, how can one keep their train of thought in the middle of researching related terms? How many pages or volumes must be on hand in order to have the entire thesaurus within reach all at once?

As for searching and browsing multilingual thesauri, there is always going to be one language whose terms that will have to carry many from the others in equivalence, because some cultures just don't have that many words for certain concept. For example: the Japanese word 'aoi' can be used to indicate a variety of blues and greens, whereas English has at least a dozen for each of those colors. This means a lot of scope notes, USE/UFs, hierarchy levels, and mappings to coordinate equivalences.

Bonus: A minute ago, I was chatting with a Belgian friend about his date and the state of his sprained shoulder earlier today, and he used the word 'reacharound'. Before I could point out anything was amiss, he said he urban dictionaried the term and clarified that was *not* what he meant. So yeah, language and cultural differences approaching syntax. =D