# Introduction to the Thematic Category Set for HT v4 (07/05/2015)

The thematic category set has been created with the intention that it allow users of the tagger to search for concepts that are of interest to them without having to look at the full Historical Thesaurus (HT) hierarchy. The result is a list of thematic category headings (aka 'SAMUELS headings').

*Creation of the Thematic Category Set*

The thematic headings list was created by looking through the category and sub-category headings in the HT. Headings which were deemed 'human-scale' have been kept, whilst those which seemed either too specific or too general were removed - for example, HT heading 01.03.01.05.12: 'Disorders of birds' (as part of the section on animal health) was thought too specific and specialist a topic for users to be likely to want to search for it; alternatively, 01.05.11.02: 'General parts' (i.e. of animals) appeared too general to be useful, although headings which are nested under it in the HT hierarchy (e.g. 01.05.11.02.04: 'Covering/skin') were considered significant enough as concepts to be given a thematic heading.

HT categories which were too miscellaneous to act as useful search terms have also been omitted (e.g. 03.11.11.42.05.08: 'Other parts' (i.e. of machines)). No HT sub-category headings have been included in the final list of thematic category headings.

Users wishing to search for categories not in the thematic headings are still be able to do so using the HT category. The thematic category/SAMUELS headings just act as shortcuts to subjects which are likely to be of interest.

*Structure of Thematic Headings*

The thematic category headings have a five-level hierarchy in the form: two upper case letters, number, lower case letter, number, lower case letter - e.g. BK01d04a: 'Board-game' (BK = leisure, BK01 = Amusement/entertainment, BK01d = A specific form of amusement/a pastime, BK01d04 = Game). For most sections of the thematic hierarchy, three or fewer of these levels are actually used, in the hope of keeping the structure as easy for a user to navigate as possible.

Each level of the thematic headings applies to all those which exist in more specific groupings below it so that BK01 'Amusement/entertainment' is considered to contain BK01d 'A specific form of entertainment/a pastime' and BK01d04 'Game'.

In the spreadsheet, the HT number is contained in the columns headed 't1' to 't7'. The thematic category alphanumeric label is given in the 's1' to 's5' columns. The thematic headings are given separately to the Thesaurus headings, although in most cases these are the same.

Where an HT category does not have a corresponding thematic dataset category, this means that all entries within the HT category (and its sub-categories) are considered to be part of the last thematic category which does exist. For example, HT category 03.13.01.05.02.01.03: 'Games similar to draughts' does not have its own thematic heading, but is part of BK01d04b: 'Board-game', which corresponds to HT 03.13.01.05.02.01.

*Other Notes*

An intentional by-product of the thematic category set is that it draws a line through the HT hierarchy indicating the point at which concepts are likely to become too narrow and specific for most humans to have much investment in learning the vocabulary associated with them.

The list of headings as they are presented here omit parts of speech other than nouns, although all parts of speech which exist for an HT heading should also be considered to be part of the thematic heading. The HT category numbers given in the spreadsheet take into account the changes to numbering which removed '00' numbers (i.e. HT v2 to v3 changes) and the restructuring of some sections especially in '01 The External World' (i.e. HT v3 to v4 changes). These changes do not affect the structure of the HT entries themselves, and each heading's primary-key value will remain the same. On account of these changes, however, some of the primary keys now appear to be out of sequence with those of the headings around them. More information on these changes can be found on the website for the Historical Thesaurus:
http://historicalthesaurus.arts.gla.ac.uk/versions-and-changes/

As a reference aid, a column containing a concatenated form of the version 1 codes has been included in the spreadsheet. In cases where the numbering has changed in the manner described above, these concatenated codes will not match the current heading number. They may be useful, however, when working with the paper version of the Thesaurus, or sources which refer to the older version of the numbering system.