

University of Debrecen  
Department of Computer Graphics and  
Library and Information Science

Information Retrieval Thesaurus  
Phonetics, Phonology and Morphology

Supervisor:  
Daniel Benediktsson

By:  
Esther Patocs

# Table of Contents

## I. Part 1

1. Introduction.....	1
2. Choice of Theme .....	2
3. Construction Policy .....	3
4. Problematic Cases.....	4

## II. Part 2

Alphabetical Thesaurus.....	7
Hierarchical List.....	28
Graphic Section	

## Sources

## 1. Introduction

The information retrieval thesaurus is a good tool in the library users' hand because it helps the orientation among indexing terminology. An average user faces great difficulties when tries to construct a query focusing on the desired topic, s/he builds up the query from their own vocabulary or tries to specify the topic but usually this is not correct enough for a machine to cope with (if we are talking about computerized catalogue), and because of this, the user gets such results that are totally different from what s/he asked (at least it seems as if this would be the case, but in effect the results differ from not what s/he asked, but from what s/he wanted to ask).

The information retrieval thesaurus provides not just a simple list of the terminology or/and a short definition of the terms contained, but shows possible synonyms leading toward the one to use, reveals hierarchic relations and other (cause – effect, point of view, process – result, process – tool, etc.) relations among the selected terms of a specialized area. The italicized phrase includes four important points in connection with thesaurus construction; let us examine these points from the back. The word area has been chosen because a thesaurus can deal not only with sciences and sophisticated, 'serious' learning but with any aspects of life, from fishing to sewing (however, people practising these claim that their hobby is a kind of art). This area should be specialized, it should use the terminology of a properly narrowed down field. It would be an extremely exciting task to excavate existing, hidden, formerly not known or not recognized relations between disciplines or various fields of life, but anyone who have made a thesaurus knows that efforts can be made, but the possibility of realization is as high as that of the great (eternal) dream of world library. The terms are not exclusively technical terms -though those form the basis- , but words used by lay people are also included as non-descriptors providing entry into the thesaurus and a path to the correct term. The selection is the only subjective factor in this work. It reflects the compiler's decisions about the necessity of elements of the word stock and above all the goal of the construction.

My objective with this work was to explore and make clear the complex relations within and among English Phonetics, Phonology and Morphology in such a way that any person not acquainted with the terminology of these fields could get a glimpse of the structure of these areas of linguistics. Though this goal at first sight does not mirror any connection with the library application of information retrieval thesaurus, but if we consider that when the user is familiar with the terminology of the area in which s/he tries to search the time of the search is shortened and the user gathers some pieces of information about the target topic without even getting the desired document, moreover, without even constructing a query with which s/he could make a request at all. Because of this the information retrieval thesaurus possesses a unique place among the reference tools of a library that helps the execution of one of the major functions of a modern library: the information provision.

## 2. Choice of Theme

The topic has been chosen according to personal consideration and experience. Being an English major student as well, I have had the opportunity to get acquainted with the major fields of linguistics, namely, Phonology, Morphology, Syntax and Semantics (for the reason of classification and the differentiation between Phonetics and Phonology see 'Problematic Cases'). All of them are about interesting topics but in terms of terminology I had greater difficulty in the case of the first two: Phonology and Morphology. In the first year of my university studies I had a course on English Phonology during which each student had to acquire the basic terminology of this field. We learned about e.g. the production of speech sounds, the position of the tongue during the production of vowels, etc. Difficulties arose when I had to report about my knowledge in the form of a term paper, because I could not organize the word stock in my mind. Finally, I chose the construction principles of a thesaurus without being aware of that: I made graphs connecting and representing the hierarchic system of the terms. Later, I did the same with the technical terms of Morphology. On the basis of this, as I am sure that I am not the only person who chose this method, I claim that not only the alphabetical section but the graphs of a thesaurus can be help in learning / searching.

However, beside Phonology and Morphology, terms borrowed from the area of historical linguistics also can be found in this work. The reason behind this was that the close relationship between phonology and morphology can be illustrated the best by historical changes in a language. As a proof of this I would like to use a concrete example: in the Proto Germanic word *stanas* [sta:nas] the root was *stan*, the *-s* was an inflection, and the *-a-* was a stem-forming suffix (*stanas* is the plural form of *stan* ('stone') in nominative case). The position of the stress was fixed, it was on the first syllable, and as language users paid less attention to its pronunciation the vowel in the unstressed syllable began to be weakened. This resulted in a reduced vowel (like the modern schwa) which later in the Old English period was completely deleted. In this way a change in pronunciation affected not just the phonological system, but the morphological system as well, because suffixes in unstressed syllables began to disappear (later this led to a change in syntax, because there were no markers referring to the words' role in a sentence, and a new governing principle emerged: word order).

### 3. Construction Policy

Beside the general rules I have been trying to apply reasonable principles during the construction of this thesaurus. As the structure of the alphabetical section and the hierarchical list is determined, creativity can work just in the graphic part. My offer is to represent the interrelations among the terms by the graphs even at the cost of squeezing too much information into one page.

First, the relationship between consonants and the places where they are produced have been indicated with special care. Consonants classified according to their place of articulation bear the name of the articulator in their own name: velars are related to the velum, dentals to the teeth, pharyngeals to the pharynx, etc. For this representation I have had to change the logical indication of hierarchic relation in the case of the organs of production of speech sounds and the subordinate terms of articulators (which is one of the subordinate terms of the former), so in these cases the broader term is located on a lower level than the narrower term, and the arrows point down toward the broader term instead of pointing upward. Of course the heads of the arrows indicate the nature of relation to dispell

confusion. Although, not all consonants are connected to an articulator or an organ e.g. labiovelars, labiodentals, palato-alveolars. These are produced by the cooperation of two articulators (labiovelars: lips+velum), and as the consonants produced by a single articulator are connected to the place of articulation, I felt no need of indicating twice the role that an organ plays in sound production. Two types of consonant could seem to be exceptions (interdentals, bilabials), but their elimination is not arbitrary, they are produced by “pair organs”, that is by one organ.

The other significant advantage of thesaurus and graphic representation that a whole process can be described by relations. By ‘whole’ I mean the starting point, the process itself, and the end point. An excellent example is the process of palatalization. This is an assimilative process when the tongue moves towards the palate during the production of velar or alveolar consonants. So it is similar to the sequential phonetic change called assimilation (a change as a result of which two sounds become more similar to each other), it occurs during the production of velar and alveolar consonants in conditioning environment (before or after a front vowel), and sounds produced in this way become more similar to palatal consonants; they are softened. Palatalization can lead to affrication: from the softened consonant a qualitatively new consonant emerges called affricate. All the italicized words are represented in this thesaurus and they are connected to each other reflecting their relationship.

#### 4. Problematic Cases

##### *Linguistics*

One word has secure position in this work, and that word is ‘linguistics’. But the revelation of its relation to phonetics, phonology and morphology was harder than I thought. The classification seemed to be easy: phonology, morphology, syntax and semantics. But what about phonetics? And what about sociolinguistics, computational linguistics and psycholinguistics? And if I say contrastive linguistics? On the basis of their names these are certain kinds of linguistics, so they should be the NT relations of linguistics. But then morphology and sociolinguistics are on the same level? Certainly not.

Finally, words representing different aspects of or different approaches toward linguistics become RT terms: contrastive and comparative linguistics, descriptive linguistics, historical linguistics, etc. Areas like sociolinguistics and psycholinguistics can be found as the subordinate terms of applied linguistics. Applied linguistics is rather an umbrella term than a proper superordinate term, but linguists use it like that, though emphasize its true nature, that is why I applied this classification. The NT terms of linguistics were determined according to the level on which they deal with the language, so they are the formerly named terms: phonology (sounds), morphology (morphemes), syntax (sentences), semantics (meaning). But in this way the question of the classification of phonetics remains open.

### *Phonetics-Phonology*

The classification of these two differs almost in every book on the topic. The differentiation does not cause difficulty, usually phonetics is defined as a branch of linguistics that examines the inventory and structure of the sounds of language, and phonology deals with the sound system (and the emphasis is on the word 'system'). But both of them can not be on the same level, according to the logic of a thesaurus despite the fact that in the literature they are and sometimes they are even listed with just a dash separating them (Phonetics / Phonology). It is certain that they are not the same, and neither can be classified as the NT of the other.

First I tried to make an RT of linguistics from phonetics, but then it did not reflect its close relation with phonology, and it was weird to see phonetics in the same role in relation to linguistics as e.g. contrastive linguistics. There is no cover term or umbrella term under which phonology and phonetics can be listed. Another problem was that the three linguistic changes - phonetic changes, phonological changes, morphological changes - borrowed from the area of historical linguistics also makes clear differentiation between the two fields (the position of the historical changes, and therefore the justification for their presence within the thesaurus is indicated on the graph that illustrates the relations of linguistics). Phonetic changes affect individual sounds, phonological changes affect the system of sounds. It is also important to note that phonetics deals with actual, individual sounds, while the concern of phonology is an abstraction. The solution is offer is to list phonetics among the RT relations of phonology, because in this way it is under linguistics, it is closely related to phonology,

but the idea that they are the same is excluded, and finally, confusion does not arise from its relation to the different approaches (e.g. contrastive linguistics, descriptive linguistics). With this classification the representation of historical changes could be solved easily on the graph mentioned previously.

### *Zero morpheme*

According to the principle that this work follows all nouns are represented in their plural forms. Processes like derivation, subtraction are in the singular, naturally. However, there is an exception that should be mentioned: 'zero morpheme'. When there is a change in meaning or addition of grammatical change to a word, and this change is not marked by an actual morpheme or change in the spelling of the word, it is called zero derivation, the addition of zero morpheme. For example the plural form of the noun 'sheep' is 'sheep', and though it is taught in foreign language classrooms that the plural of 'sheep' is the same as the singular, for linguists this is not true. When we try to express something different we never use the same word, and the plural of a noun is certainly a different thing than its singular. On the basis of this, linguists claim that the plural form is not the same because there is a morpheme attached to the word, just it does not have an actual form.

sheep (root) SING + 0 (zero morpheme) = sheep (derived word) PL

It would not be logical to use the plural form of a word that does not have a form at all. 'Zero morphemes' would sound strange, it suggests that there can be various forms in which zero morpheme can appear, while it appears in one single form: when it is not indicated.