Institutionen för Lingvistik
Stockholms Universitet

Vt-98

The Rhyme-matcher

Tomas Emmer, Martin Hassel

The Rhyme-matcher is a computer program designed for locating the last stressed syllable, which is essential for rhyme matching. It does so by assigning stress to the appropriate syllable. The Rhyme-matcher is also programmed to distinguish the orthographical pattern of the Swedish simple word. It allows or disallows input on the basis of this. Testing has shown reliability in getting output. Output analysis and further testing remains.

Påbyggnadskurs i datorlingvistik
Uppsats 10 poäng
Handledare: Gunnel Källgren

Contents

The Rhyme-matcher

1 Introduction

Rhyme is a poetical style of writing where two or more nearby words or syllables sound alike. The purpose of the rhyme is to combine verses to higher units (Oldberg 1945). If the conforming sounds are in the beginning of the words, we call it alliteration; *manisk måndag*, 'manic Monday'. Assonance consists of conformity between two sounds in the middle of the words; *garn*, 'yarn', *horn*, 'horn' and *ful*, 'ugly', *hus*, 'house'. The end rhyme demands conformity between the ending sounds from a specific point in each word. Alliteration and assonance can with a unitary term be called ancient rhymes (fornrim; Beckman) and are common in ancient Nordic poetry. The Vikings, Celts, ancient Italians and Teutons, especially, frequently used alliteration. When taking a closer look at the old Swedish laws you discover they are commonly written in alliteration; *Land skall med lag byggas*, 'Land shall with law be built', and both alliteration and assonance are common in sayings and proverbs and, not in the least, advertising; *råg i ryggen*, '-', *summan av kardemumman*, 'the sum of it all', *frisk och rask*, 'strong and healthy', *bakar bättre bröd*, 'bakes better bread'.

In common speech the term rhyme almost always denotes the end rhyme. The end rhyme, which we will concentrate on in this paper, is most likely an elaboration of alliteration and assonance and has during the last 700-800 years stepped into a position of dominance. It was the poets of the church who took the end rhyme to their hearts and incorporated it into their sermons as a method of memorisation and rythmisation. The end rhyme is often found in proverbs; *många bäckar små bildar en stor å*, 'many a little makes a mickle', and it is here it first becomes common. It can be traced as far back as to the ancient runic writings but is there by far overweighed by the alliteration.

1.1 Purpose

The purpose of the Rhyme-matcher is to take a given word or non-word[1] e.g *spruts*, basically an string of graphemes, and either try to find a matching rhyme, or rhymes, in a lexicon or to test it against another string of graphemes in order to conclude if the two strings, words, rhyme or not. The idea is that this piece of code could be integrated in a title developed by ELD (see below) and published by Levande Böcker to enable children to recognise the structure of (written) words in a lustful manner by playing with rhymes. One of the crucial steps of acquiring a language is play. The remodelling of and playing around with the acquired language competence is the child's way of analysing and understanding. It is through carefully observing the language competence of small children, that are still in the learning process, that we learn that their learning is not robotic imitation, but rather creative learning (Söderbergh 1985 p 89ff).

1.2 Background

This paper is written in conjunction with practice. Much of the work on the Rhyme-matcher has been conducted in the premise of Levande Böcker and ELD´s facilities. All of the coding process has been done on computers made available by ELD.

Levande Böcker is a publishing company. They publish interactive educational computer programs in Swedish. ELD is a production company that work in close relation with Levande Böcker, they produce interactive CD-ROM titles for children.

[1] Our definition of a non-word is a string of characters that isn't a lexical item but still forms a word token that adheres to the patterns of Swedish phonotactical structure.

The goal of the practice on ELD and the co-operation with ELD and Levande Böcker has been to create a program that would help children to acquaint themselves with preferably written language in a lustful manner (e-mail, 98-03-09). The program could be integrated in a title that Levande Böcker are in the process of creating.

The idea of the Rhyme-matcher came in close discussion with Jonas Beckeman at ELD and our tutor Gunnel Källgren. Jonas' experience as a programmer was of great use to us in the making of the Rhyme-matcher. We coded the program in a to us unknown computer language, Lingo. Hassel is however an experienced computer programmer and learned to master the language in a short period of time. Emmer contributed with his knowledge of children's development and his experience of working with children.

1.3 Method

To accomplish its task the Rhyme-matcher has to analyse the inputted string of graphemes' syllable structure and handle discrepancies between the orthographical representation and the phonematic reality. We have only had the time to, in reality, deal with the former part even though the latter is to some extent discussed in this paper. The syllable sub-task basically consists of analysing the inputted string of graphemes' syllable structure, treating it like a simple word, in order to find the last (main) stressed syllable. This is needed to be able to match two analysed and annotated strings of graphemes for rhyming.

2 Data

The data used in the work for this paper have been taken from the Stockholm Umeå Corpus Version 1.0, SUC 1.0, Eva Ejerhed, Umeå University, Gunnel Källgren, Stockholm, Copyright (c) 1997 Dept of Linguistics, Umeå University, and Dept of Linguistics, Stockholm University.

```
The SUC corpus was created as part of the joint research
    project "Corpus based research on models for processing
    unrestricted Swedish Text" ("Korpusbaserad utveckling av
    modeller för datoranalys av löpande svensk text") between
    the Departments of Linguistics at Stockholm University and
    Umeå University respectively. The principal investigators
    were Gunnel Källgren in Stockholm and Eva Ejerhed in Umeå.
    (CD-ROM suc 1.0)
```

We have used the raw text from the SUC to extract a word occurrence list. The raw text is annotated in SGML format for parts of speech (CD-ROM suc 1.0). The purpose of extracting a list from the SUC-corpus was to test the Rhyme-matcher. The work with the occurrence list became more difficult and more extensive than we had anticipated. The list cannot be seen as a tool for anything else but for testing the Rhyme-matcher. We operated from the Unix environment and ran the SUC raw text through the UNIX programs grep, sort and uniq. To make the handling of the list more convenient and controllable we divided the list into alphabetical disk files, consisting of words beginning with the same letter. The list consists of a file for every Swedish alphabetical letter.

We were solely interested in the simple word (for clarification of the term see section 3), but the corpus is built on running text and therefore we choose to extract every string of characters between the SGML tags. We were not interested in strings of characters that weren't words. So we removed all strings of characters that weren't letters. That meant that we removed numbers, question marks, punctuation marks and all other characters that don't belong to the Swedish alphabet. We also choose to remove names such as personal names,

names of cities and towns and so forth. We removed foreign words when we found them, however the corpus contains almost 300000 word types and we did not have the time or skill to remove them all. And as you can see in section 4 the foreign words play a vital part in demonstrating how the Rhyme-matcher works, so for educational reasons we luckily missed some.

3 The Swedish word

In our description of the Swedish word we will concentrate on the simple word. We will not give any extensive description of the structures of compounds, the Rhyme-matcher is solely meant for dealing with the simple word. However the distinction between the simple word and the compound word is complex and the distribution and combination of phonemes/graphemes are quite distinct. So obviously some clues to the simple word structure lies in the structure of compounds.

The compound words can consist of combinations of root morphemes with other root morphemes e.g. *handbok,* ‘handbook’, or root morphemes combined with affixes, *dumhet* ‘stupidity’, often as a result of derivation. Compounds such as the former consisting of two or more root morphemes that can stand-alone as words by themselves are sometimes called genuine compounds. The latter compounds that share the same suprasegmental pattern as genuine compounds such as stress, but differ in morphological pattern can be called formal compounds (Sigurd 1965 p 28).

Sigurd has constructed a model for describing the simple word where the following types of simple words can be distinguished:

1. Words consisting of only one stressed syllable (monosyllables, e.g. *häst*, ‘horse’, *hund*, ‘dog’.
2. Words consisting of a stressed syllable followed by one or several unstressed syllables e.g. *spindlar*, ‘spiders’, *överste*, ‘colonel’.
3. Words consisting of a stressed syllable preceded by one or several unstressed syllables e.g. *problem*, ‘problems’, *parad*, ‘parade’.
4. Words consisting of a stressed syllable both followed and preceded by one or several unstressed syllables e.g. *problematis*k, ‘problematic’, *karakterisera*, ‘characterise’.

Genuine Swedish simple words often follow the description of the first two types. The last two types generally represents loan words (Sigurd 1965 p29).

3.1 Phonotactic structure

One of the Rhyme-matcher’s goals is to decide whether a string of tokens, a word or a non-word is coherent with Swedish natural phonotactical pattern. Phonotactical patterns are governed by the actual strings of phonemes/graphemes that occur in the language.

Sigurd (1965) gives a model for describing the distribution of phonemes in the Swedish simple word. Sigurd’s study is based on *Svenska Akademins ordlista* (SAOL) including about 200 000 entries. In the beginning of the making of the Rhyme-matcher the intention was to give every string of tokens handled by the program a phonological/phonetical description, but restriction on time made us choose to only incorporate a description of orthographic distribution.

Sigurd says that there is a fundamental difference between description of grapheme combinations and phoneme combinations. He says that phoneme patterns must be regarded as primaries since the phonetic constitution of the phonemes are related to their distribution. We fully agree with Sigurd’s conclusions, but we also believe that the Rhyme-matcher gives valuable and often correct assumptions based on an orthographic description only. We have taken the liberty of translating Sigurd’s description of phoneme distribution in consonantal clusters and combinations with vocalic elements to orthographic distribution and combination patterns. We are aware of the fact that this translation gives the Rhyme-matcher less credibility. What we miss in the translation is the phonetic information that lies in the combination of phonemes. What we gain is time.

However, even though the distributional structure can be compared with phonetic structures on different levels (articulatory, acoustic and perceptional) and that the connection of these properties seems natural, this paper is mainly concerned with how the structure and pattern of phonemes/graphemes appear and not why. We will though describe some of the essential phonetic properties of the consonantal cluster members such as sonority value and phoneme combinability. The sonoric value of a sound has great influence on its distribution and a sound’s phonetic properties give restriction to combinability. The phonemes will be annotated between slashes // and the approximation[2] of the corresponding grapheme between ||. The correspondence between sound and writing in Swedish in not at all clear-cut and when further explanation and description of the correspondence is necessary we will try to comment on this.

The phonotactic structure of the Swedish simple word can be described as an alternation between vocalic and consonantal elements (Sigurd 1965 p 30). To describe the distribution of phonemes in the Swedish simple word you need to consider what restrictions they follow. Which phonemes combine; restrictions in membership, in which order does the phonemes appear; restriction in sequence and how long sequences occur; restriction in number of members.

Different positions of the consonantal members can be distinguished in the simple word. Initial, medial and final. Within these positions further positions can be defined (Sigurd 1965 p 43). Below follows a table that shows the phonotactic pattern of the simple word. The table also illustrates the different positions initially, medially and finally of consonantal members.

[2] There is no one to one correspondence between phonemes and graphemes in Swedish. What the authors have had to do is to approximate the correspondence regardless of the differences that different speakers might have in their speech.

C A6	V A5	C A5	V A4	C A4	V A3	C A3	V A2	C A2	V A1	C A1	V0	C P1	V P1	C P2	V P2	C P3	V P3	C P4	V P4	C P5
In 6		In 5		In 4		In 3		In 2		In 1		F 1		F 2		F 3		F 4		F 5
		M a5		M a4		M a3		M a2		M a1		M p1		M p2		M p3		M p4		
										h	a	t								
										spr	a	tt								
										b	a	d	e	rsk	o	rn	a			
	a	m	e	r	i	k	a	n	i	s	e	r	a							
								pr	o	bl	e	m								
										h	e	d	e	rl	i	g	a	st	e	
										m	ö	rkr	e	t						

Table 1. All the words are orthographically described in contrary to Sigurd's description (Sigurd 1965 p 30) which are phonematically described in the diagram. C = consonantal position, V = vocalic position, In = initial position (the position before the first vowel), F = final position (the position after the last vowel), M = medial position (intervocalic), a = before (ante) the stressed vowel (V0), p = after (post) the stressed vowel (V0), Subscript numbers denote the distance from the stressed vowel.

The initial, medial and final positions of the consonantal clusters are although connected quite distinct. As you can see in the word *problem* 'problem', we have a medial consonantal *–bl-* cluster that also can stand as initial; *blad*, 'leaf'. However, the medial cluster in *mörkret*, 'darkness', *-rkr-*, can never occur as initial, nor final. We can also see that the only initial three-membered clusters in table 1. are initiated with |s|, e.g. *spratt*, 'prank' and *skruv*, 'screw'. The two-membered clusters in the table can be initiated by either |p| or |s|, e.g. *problem,* and *spindlar*. There seems to be a number of restrictions to the distribution of the phonemes/graphemes in the Swedish simple word.

3.1.1 Initial sequences

The position immediately adjoining the vowel can be occupied by any of the consonants. This analysis shows that the further away from the vowel the member stands, the fewer members can occupy that position (Sigurd 1965 p 44). We can talk about initial position classes. No initial consonantal cluster can be longer than three members[3], no three-membered initial cluster can start with anything but /s/. The first of the position classes then only contains /s/. The second initial position can be occupied by any of the following consonants; /s/, /f/, /p/, /t/, /k/, /b/, /d/, /g/, /m/, /n/. The third position is the one adjoining the vowel. These positions have strong correlation with the theory of sonority and are phonetic properties of the members in consonantal clusters. Some sounds are inherently more sonorant (sonorance or resonance) than others with vowels as the most sonorant and obstruents as the least. The scale is a matter of degree, for example the vowel /a/ is more sonorant than the vowel /i/ or /u/ (Spencer 1996 p 89).

[3] This refers only to number of phoneme members. The number of graphemes can in loan words be more than three members e.g. *schlager* which starts with the phoneme /ʃ/.

Sonority scale for consonants:

Class	Phonemes	Value
Glides	/j, w/	5
Liquids	/r, l/	4
Nasals	/m, n, ŋ/	3
Fricatives	/v, ð, z/	2
Plosives	/p, b, t, d, k, g/	1

All the phonemes from Spencers list are not included in this list simply because we did not have the correct fonts for the ones missing. However, this scale explains a great deal of the phonological pattern of the distribution, but some phonemes do not adhere quite as neatly as the rest. For example, /s/ breaks up the sonority scale and occupies positions before less sonorant sounds e.g. *spade*, 'shuffle', where /p/ the bilabial soundless plosive stands in the position after the /s/. Apparently the sound /s/ holds a unique position phonetically and we can also see it holds a similar position orthographically. An attempt to explain the unique position of /s/ is to treat the two membered clusters it appears in e.g. *sp*, *st* and *sk* as affricates. There could be a phonetic motivation for this interpretation. The [p,t,k] phonemes lack aspiration adjoining to /s/ in an initial cluster. Standing by themselves initially they are all aspirated (Sigurd 1965 p 62).

The combinations of the members in consonantal clusters are also restricted. A list of actual occurrences of different two-membered consonantal clusters in SAOL and their member combinability is presented below. As would appear apparent /r/ and /s/ have the highest score. But it is noticeable that /r/ and /s/ do not combine with each other.

Phoneme	Combinability
r	8
s	7
l	6
j	5
v	5
n	5
k	5
p	4
f	4
t	3
b	3
g	3
m	2
d	2
h	0
ç	0
ʃ	0

As we can see from the list the phonemes /ʃ/, /ç/ and /h/ do not combine at all. They can only occur by themselves. Orthographically they correspond to a number of consonantal clusters. The /ʃ/ and /ç/ phonemes can be said to correspond to |stj|, |skj|, |sk|, |tj|, |k|, among others. The phoneme /j/ corresponds to |hj|, |lj|, |j|, and in some loan words |dj|. This is not an exact representation of which sound corresponds to which grapheme. To construct such a representation is almost impossible. This differs widely depending on the speaker's age, dialect and social-class. The approximate correspondence is meant to illustrate the number of different grapheme clusters the authors had to add on to Sigurd’s description of phoneme distribution. A further discussion of the implications of the translation of phonemes to graphemes will be presented in section 6.

3.1.2 Final sequences

The morphological structure has no influence on the structure of initial sequences in the Swedish simple word. We have no prefixes that solely contain consonants. But in the study of final and medial consonantal clusters the morphological structures must be taken into account (Sigurd 1965 p 67).

There are a number of final sequences that only occur in inflected or derived forms e.g. *-rmst* in *närmst* '‘nearest'. There are also, at least in principle, derived and inflected forms of words with final sequences with as much as 8 consonantal members.

Stem	adj.	neuter	gen	resulting cluster
Ernst	*sk*	*t*	*s*	*ernstskts*

Therefore it would be motivated to distinguish between polymorphemic and monomorphemic sequences. However the distinction between these different sequences cannot solely be based on a primary system and a secondary system on the criterion: monomorphemic – polymorphemic. Polymorphemic sequences as in *mans*, ‘man’s’ can conform to a natural phonotactic pattern (monomorphemic) as in *svans*, ‘tail'’ *dans*, ‘dance'. So polymorphism cannot be taken as evidence that the sequence in it is deviant (Sigurd 1965 p 67).

A solution could be to take ease of pronunciation as a way to decide whether a sequence is secondary or not. A sequence like *–mskt* in *hemskt*, ‘awful’, is often pronounced *–mst* with the /k/ silent. This could be regarded as evidence that the sequence is secondary. Sigurd has adopted a set of mechanical rules that exclude certain forms from the natural phonotactic pattern. The rules follow below.

1. genitive forms ending in s as *marschs*, ‘of the march’
2. adverbial forms ending in *s* as (till) *lags*, ‘satisfactory’
3. passive forms ending in *s* as *följs*, ‘is followed’
4. neuter, adverbial and supine forms ending in t as *skev*t, ‘wry’, *ärvt*, ‘inherited’
5. participial forms ending in *d* as *ärvd*, ‘inherited’
6. superlative forms ending in *st* as *närmst*, ‘nearest’
7. accidental derivations ending in *sk* as *skälms*k, ‘roughish’

As we see from these rules *s* also has a unique position in final clusters. It breaks the natural phonotactic pattern and is in the first three and the last two rules regarded as secondary. Sigurd has no discussion concerning *s* in the final clusters. To treat it as an affricate in the final position would be quite wrong while the following [p, t, k] sounds as in *skälmsk* and *närmst* clearly are aspirated. The authors has no alternate solution to the to the problem, we can but state the fact that *s* does not follow natural phonotactic patterns.

3.1.3 Comparison between initial and final sequences

If we compare the structures of final and initial consonant clusters, we find a variety of interesting differences. The following observation concerns two-membered sequences. The number of permitted final sequences is much greater than the number of permitted initial sequences. The members of the final sequences more freely combine than the initial do.

Phonetically related consonant sounds as stops [p, t, k], nasals [m, n] liquids [r, l] and labials [m, b], among other combinations, combine finally. These sequences are not permitted initially. Inversible clusters are only found in final position; *st*:*ts*, *sp*:*ps*, *sk*:*k*s. The number of permitted three-membered clusters are also much greater than the initial ones (Sigurd 1965 p 106).

3.2 The syllable

The end rhyme demands conformity between all sounds in the two rhyming words from the point of the vowel in the last stressed syllable to the end of the words; *fet*, 'fat' – *het*, 'hot', *falla*, 'to fall' – *tralla*, 'trolley'. In order to find the last stressed syllable in a word you first have to identify the syllable boundaries. There has been much debate on how and if this can be done (and if it is meaningful at all to even discuss the term 'syllable' (Dahlerus 1994 p 2)). A number of alternate models of the syllable have been proposed. Going through and comparing them is beyond the scope of this paper. The starting-point for this is that the lexical units - morphemes and words - in their sound representation not only consist of an array of speech sounds - consonants and vowels - but that these also are grouped in a characteristic way into syllables. The syllables psychological reality as a relevant unit in the sound structure of a language should be evident, for instance, from the studies of children’s language and speech development. It has been shown that the syllable is a reality for the child even before the structure in consonants and vowels is achieved (Bruce 1998, Lexikalisk prosodi p 2).

3.2.1 Syllable structure

Many phonologists envisage a *branching, hierarchical* syllable structure. For the traditional structuralist approach see Pike (1967) and Pulgram (1970). More recently, writers like Kiparsky (1979), Halle and Vergnaud (1980), Steriade (1982) and Harris (1983) have also presented a reworked version of the hierarchical branching theory, presenting a multi-tiered phonological theory (Katamba 1989 p 153-154). In this paper we will concentrate on the branching hierarchical syllable structure view.

A. Noreen (1907) states that the syllable boundary should be placed in an intensity or sonority minimum somewhere between two peaks of sonority (usually, but not always, vowels), and that we can freely choose where this (intensity) minimum should be placed as long as the word is naturally pronounceable to a native speaker of the tongue in question. This intensity minimum should preferably be placed in conjunction with morphomatic boundaries (Dahlerus 1994 p 3). If analysing the syllables from a *sonority scale* of view you soon discover that the lesser sonoric sounds group themselves around the most sonoric sounds in a gradual manner. This way we can discern a pattern of plausible consonant clusters, plausible because only some are allowed before the sonority peak, some after and yet some are allowed both before and after the peak.

If we take a closer look at the syllable with the end rhyme in mind we can't fail to see that it is the final part of the syllable that is important here. Even very young children recognise that words like *katt*, 'cat', *natt*, 'night', *matt*, 'weak', *(få) fnatt*, '(go) bananas', *satt*, 'sat', sound alike in some manner. Dissecting a simple rhyme as *katt*, 'cat' – *natt*, 'night' we see that the common part is |-att|. Thus, we can split up the words in to |k-att| and |n-att|. The second part is called **rhyme**, since it is it that is doing the rhyming, and the first part is called **onset**. The onset, when present, is a string of one or more consonants. When the onset is the common part we call it alliteration. All syllables have to have a peak, sonoric nucleus, which, as mentioned above, usually is a vowel (or a diphthong). This peak, or head, is the only obligatory part of the syllable (all languages taken in consideration). Where a syllable ends in one or more consonants we can divide the rhyme into two distinct parts; the **nucleus** (the syllable head) and the **coda** (the ending consonants, also called tail or margin). In this case the syllable is said to be closed. When the coda is missing the syllable is said to be open. Some languages only permit open syllables but even in those languages that allow closed syllables, there is often a clear preference for open syllables. We now have the following structure; (C*)-V(V)-(C*) or (onset)-nucleus-(coda), with a preference for onsets before codas. Representing this in a branching, hierarchical structure this would look like:

```
           Syllable
          /        \
   (Onset)          Rhyme
                   /      \
           Nucleus         (Coda)
```

What we now need is a theory that provides us with a way of grouping arrays of CV elements into syllables when encountering situations like this; VCVCCCVC. From what we have established above it is clear that each V-element will be associated with a syllable peak. What we now need to show is to which syllable peak C-elements are to be assigned in ambiguous cases, where they could go with either the following or the preceding vowel. Katamba (1986) refers to (Kahn 1976, Clements and Keyser 1983) about the **Onset First Principle** that has been proposed to deal with such cases:

[9.9] (a) 'Syllable-initial consonants are maximised to the extent consistent with the syllable structure conditions of the language in question.'

(b) 'Subsequently, syllable-finial consonants are maximised to the extent consistent with the syllable structure of the language in question.'

(Clements and Keyser 1983:37)

Principle (a) applies before (b) in any derivation, i.e. in potentially ambiguous cases initial consonant clusters take precedence over syllable final ones.

In text en clair, unless there is a overriding language-specific reason for doing otherwise, given a string like VCV, the Onset First Principle requires that the string be divided up as V-CV rather than VC-V. Now all we need to know is which consonant clusters are allowed in Swedish and we can even handle complex words like *vitaminspruta*, 'vitamin shot' which is divided up to CV-CV-CVC-CCCV-CV, although the Rhyme-matcher is, at this stage at least, solely constructed to handle the simple word, not compounds like *vitaminspruta*.

An example of a language-specific reason in Swedish is the CC-sequence |sp| that is allowed both initially, as in *spara*, 'save' and finally, as in *rasp*, 'grater', so in the case *raspa*, 'grate' it could either be divided up as |ra-spa|, |rasp-a| or |ras-pa|. The Onset First Principle predicts the former, CV-CCV, to be the correct syllabification even if the one syllable in the stem, |rasp|, is formed like CVCC.

When two vowels meet, so called hiatus, the consensus is that the syllable boundary is to be placed between the two vowels, for example *video*, *kaotisk*, *februari*, while the sequence [au] in *pausa* and *faun* form a diphthong and belong to the same syllable. We neither have the need, the time nor skill to create a program that makes a complete syllable structure analysis so we have decided that we can handle the diphthong and the hiatus in the same way without any lack in performance. We simply handle all vowel sequences as hiatuses, a syllable sequences.

Setting up clear-cut rules for discerning syllables is, as mentioned, a hard nut to crack. And as mentioned, maybe it's not always necessary to make an unambiguous delimitation between syllables. Partly the question of where the syllable boundaries are drawn up is to the individual speaker's intuition. The problem is that it's very difficult to build intuition into a computer program. Only to avoid misunderstandings, maybe it is best to mention that the phonological syllable delimitation we discuss here is not to be confused with the prevalent rules of syllabication of written words officially agreed upon.

3.2.2 Syllable weight

One distinction we soon discover to be very important is difference in *syllable weight*. The consensus today is that more important than the traditional classification of phonological systems in terms of open and closed syllables is their classification in terms of weight. In many languages a factor that determines the applicability of certain phonological rules is the weight of the rhyme. Many languages distinguish between two types of syllables, *light* and *heavy*. Basically, a syllable is light if it contains a nonbranching rhyme and heavy if it contains a branching. This is called the **Branching Rhyme Hypothesis**. The essence of this hypothesis is to locate branching anywhere within the domain of the rhyme, be it at the level of the nucleus and coda, or within the nucleus itself. The onset never seems to play any role, in any language, in the computation of syllable weight and, consequently, its internal structure is irrelevant. In Swedish the distinction between light and heavy syllables is drawn in the following way; in a light syllable the rhyme contains a short vowel, (C*)V, and in a heavy syllable the rhyme contains either (i) a long vowel or diphthong optionally followed by one or more consonants, (C*)VV(C*), or (ii) a short vowel followed by at least one consonant, (C*)VC(C*). Some dialects permit stand-alone diphthongs (and even triphthongs!!!), without these counting as heavy, like the southern Swedish /ei/ but these are not written as two following vowels in written language (other than when pointing out that someone is speaking this particular dialect).

3.3 Stress

A syllable is either stressed or non-stressed. Stress is primarily a matter of greater auditory prominence. An element that is stressed is highlighted so that it becomes auditory more salient than the rest of the elements in the string of which it is part. The main phonetic ingredients of stress are *pitch*, *length* and *loudness*, but loudness is a much less important parameter than pitch or length. A stressed syllable is pronounced with more energy and is longer than it's non-stressed counterpart. It is a great difference between a stressed and a non-stressed syllable. According to Håkansson and Stenquist (1989 p 15 our translation) the following applies to stressed respectively unstressed syllables:

A stressed syllable	A non-stressed syllable on the other hand
- is long	- is short
- is strong and has extra pressure	- is weak
- has any of the 22 vowel sounds	- does not make any difference between long and short vowel sounds
- has melodic variation	- is monotonous

They then continue by saying that there are two types of stressed syllables (in Swedish, our note):

- Long vowel sound followed by none or a short consonant sound
- Short vowel sound followed by a long consonant sound

Unfortunately they conclude by saying; A word (in Swedish, our note) has one or two stressed syllables never more. If a word only has one syllable, it is stressed. In words with two or more syllables, you can not know which syllable is stressed and which is not. Stress has no fixed place in Swedish. When learning new words, you therefore always have to learn how they are pronounced. Fortunately, others disagree.

Stress is not predictable from the syllabic structure alone. A disyllabic formative of the form /CVCVC/ can have stress on the first or the last syllable; 'lâkan, ba'nân, 'pajas, fa'tâl, ... However, if the tenseness of the vowels is also taken into consideration it is possible to predict stress by rule for the majority of Swedish words. Unfortunately we cannot predict the tenseness of a vowel just by looking at it and its context, i.e. we cannot discern whether a vowel is long or short and thereby the whole approach seems to fail.

Fortunately, again, this is not the whole truth about stress. Every lexical item has one primary stress (Malmberg, 1956, p. 101; Elert, 1964, p. 16). There are two fundamental stress patterns in Swedish, namely *simple* and *compound*. The simple stress pattern is characterised by one stress (main stress) that is placed on one of the three last syllables in the stem of the word. Sometimes, especially in long words, one may notice a secondary stress in the beginning of the word. This extra prominence is not to be considered as a genuine stress but rather as a rythmic induced secondary stress belonging to the phrase. The compound pattern consists of two genuine stresses: one main stress in the beginning of the word, for example in the first part of a compound word, and a secondary stress late in the word, on the last part of compound word. When a compound word is formed out of more than two parts the medial part receive no stress. The fundamental systematics of placing stress (main stress as well as secondary stress) in the individual parts is the same as in noncompound words. Monosyllabic lexical items, for obvious reasons, only have one strong stress. In polysyllabic words the primary stress may fall (Lindau 1970 p 2-3):

1) on the first syllable of the word in
   a) nouns ending in vowels
   b) nouns ending in -el, -er, -en
   c) nouns ending in other vowel-consonant combinations (*-on, -op, -or, -ott, -an, -ad, -ap, -ak, -as, -us, -ud, -ul, -um,* etc.)
   d) verbs
   e) other lexical categories
   f) nouns with more than two syllables
   g) in words with certain prefixes, including *bi-*, *an-*, *om-*, *in-*, *av-*, *upp-*, *ut-*, *vid-*, *till-*, *väl-*, *för-*, perhaps *själv-*
2) on the syllable following certain other, unstressed, prefixed, including *be-*, *ge-*, *ent-*, *i-*. *för-* and *väl-* resides in this group as well
3) on the last syllable in a great number of words
   a) ending in a vowel
   b) ending in consonant(s)
4) on other syllables

The first thing one may notice is that stress is not an inherent vowel feature. It is an autosegmental property of the word. Its location in the phonetic representation of a word may depend on the presence of certain affixes or grammatical information such as whether the word is realised as a noun or a verb. First, there are affixes whose presence has no effect on the primary stress of the root to which they are attached. Another class of affixes attract stress to themselves as though they were magnets. When they are attached to a word they always get the main stress. If first considering these, we can lessen our burden in the hunt for the stressed syllable.

3.3.1 Morphology and stress

Swedish has a rather extensive morphology with great opportunities to combine words and to derive and inflect them. The root is the simplest possible form of a lexical item, upon which all other bound and free forms involving that morpheme is based (Trask 1996 p 244). The stem is a bound form of a lexical item which typically consist of a root to which one or more morphological formatives have been added and which serves as a base for the formation of some further form or forms (Trask 1996 p 259ff). Affixes carry information about stress, some are inherently stressed and others not. We include both types in the database.

The process of word-formation is in principle based on the conception of morphemes. The noun *manlig*, 'manly', consist of the free morpheme *man-* 'man' which is a lexical item and the bound morpheme *-ig* which is an adjective derivational suffix. A free morpheme can generally stand by itself as a word, a bound morpheme can not. Derivational and inflectional morphemes are usually bound morphemes.

Swedish derivational morphology is extremely productive and consists of affixes that derive words from one kind of part of speech to another or modifies the meaning in the same category. The inflectional morphology is more fixed and not as productive. We will include a description of Swedish inflections for nouns, adjectives, and verbs.

3.3.2 Prefix derivation

The prefix-derived words can be divided into two main groups;

1. Derived words with prefixes with main or secondary stress
2. Derived words with prefixes with weak stress

The first group consists of nouns, adjectives and verbs e.g. *miss/tänkt*, 'suspicious' or 'suspect', *pre/destination*, 'predestination'. These prefixes can be domestic or foreign. The second consist solely of verbs made from the prefixes *be-* and *för-* which are loans from German e.g. *be/tänka*, 'consider', *för/undran,* 'amazement'.

A group separate from these are the words derived with the German prefixes *an-*, *bi-*, and *und-*, these often make verbs e.g. *an/föra*, 'lead' or 'command' (Thorell 1981 p 60).

Derivational prefixes with strong stress are often foreign, mainly Greek or Latin but there are domestic prefixes with strong stress such as o- e.g. *o/svensk,* 'unswedish'.

3.3.3 Suffix derivation

The suffix derivation in Swedish is more extensive than the prefix derivation. The structure of the base, that means the simple, compounded or derived word that the derivation is made from, can be represented as below.

1. The base is a simple (monomorphemic) word e.g. the noun *man* ‘man’, in *man/skap*, ‘crew’.
2. The base is a compound e.g. the noun *sensommar*, ‘late summer’ in the adjective *sensommar/aktig*, ‘late summerish’.
3. The base is a derived word e.g. the noun *sten*, ’stone’, in the adjective *stenig*, ‘stony’ in the noun *sten/ig/het*, ‘stoniness’

The base can be polymorphemic, a base consisting of more than one morpheme, and more than one affix can be added e.g. *o/egent/lig/het/er/na/s*, ‘the improprieties’. This word consists of 7 morphemes, 1 root, 1 prefix and 5 suffixes. This is not uncommon in Swedish word-formation. It is in this case hard to say if it is –*egent*- that is the root. Intuitively -*egentlig*- would be the root, perhaps lexicalized but –*lig*- is an adjective derivational suffix and could be said to have been added. We won’t go deeper in this analysis, the point is that polymorphemic bases like this one, must be handled by the Rhyme-matcher. Further reading about how will be presented in section 4.

We could divide the types of derivation into:

1. Derivation of part of speech e.g. *vänskap/lig*, ‘friendly’
2. Derivation with zero-suffix e.g. *glid* (:*glida*), ‘glide’ (to glide)
3. Regressive derivation e.g. *vinterbada* (:*vinterbad/are*), ‘to bath in the winter’, (a person who baths in the winter) (Thorell 1981 p 70)

We are solely interested in the first group. One can further divide the first group into two separate systems for derivation. They both have foreign origin, one from nordic-germanic and the other from the romanic, mainly Greek, Latin and French. The both systems differ in the stress-pattern. Adjectives like *ivr/ig* ‘anxious’ and *verk/sam*, ‘effective’ or ‘energetic’, can represent the nordic-germanic system. The adjective *effekt/iv*, ‘effective’, represents the romanic. The type *ivr/ig* has grave accent with main stress on the base and weak stress on the suffix –*ig*. The type *effekt/iv* has acute accent with main stress on the suffix –*iv*. As romanic derivations regularly have stress on the suffixes, the main stress often will be on a syllable in the derivation rather than a syllable in the base.

When we have only what we think is the stem, or as little as possible but the stem, left, we can place the main stress on the first heavy syllable, counting from the right-hand end of the word. Where disyllabic words contain no heavy syllables, stress falls on the first syllable from the left. This mainly to make our quest a bit easier. It is true that it is a common idea that Swedish in reality has its main stress on the first syllable but a comparison with English clearly shows that English has a much stronger preference for initial main stress. The obvious finding is that main stress often is placed later in Swedish than in the corresponding word in English.

There is also a number of words where placing the main stress becomes difficult. This is usually in the case of foreign words, not in the least foreign names. When these words and names are to be adapted to the Swedish phonotactics, situations can arise where there is great uncertainty about where the main stress should or should not be placed. Some good examples of this are *Canberra*, *Monaco*, *Adidas*, *oboe*, *omega* and *viola*.

3.3.4 Gemination

What is also to be noticed is that orthographically geminated consonants in Swedish give away stress. In Swedish a consonant gets geminated if it is located after a stressed vowel and one of the two following conditions is fulfilled:

1) The consonant precedes a morpheme boundary as in *kall*, ‘cold’ (adjective) or *kall-t*, ‘cold’ (adverb)
2) The consonant precedes a vowel as in *Kalle*, a common Swedish name

In the word *salt*, ‘salt’, for example, the consonant following the stressed vowel precedes another consonant without there being a morpheme boundary between them and, hence, there occurs no gemination. What we have done is to assume that when a gemination is found, the preceding vowel is stressed. Unfortunately we haven’t had the time to implement geminated consonants followed by one or more, other/different, consonants. We’ve only had the time to implement geminated consonants followed by one or more vowels. Some foreign words also defy these rules. Typical problem words are *parallell*, ‘parallel’, and *cigarrett*, ‘cigarette’.

3.3.5 Minimal pairs

This is a problem that arises is Swedish since it uses (main) stress placement as a distinctive feature. It should be noted that both the vowels and the consonants quality is influenced by stress (and nonstress) in such a way that it sometimes can be hard to vindicate that stress is the only difference. However, it is still a problem, especially since the quality of the vowels and consonants aren't even considered by the Rhyme-matcher. Some minimal disyllablic pairs, inflected and uninflected forms mixed, are given below for reference.

1	2
'banan	ba'nan
'modern	mo'dern
'varan	va'ran
'fasan	fa'san
'finnes	fi'ness
'syntes	syn'tes
'trumpet	trum'pet
'kantat	kan'tat

4 Algorithm / The machine

1) **analyseraOrd()**. Analyse inputted string of graphemes.

1.1) **kollaUttal()**. Check if the inputted string of graphemes is pronounceable using the Onset First Principle.

1.1.1) Check for initial consonant cluster, if found, verify that it's a legal one by comparison with partial strings in predefined lists.

1.1.2) Go through the rest of the consonant clusters, one by one (loop).

1.1.2.1) Get next consonant cluster and verify that it is a legal final cluster by comparison with partial strings in predefined lists, final for efficiency. (See discussion in section 6.3).

1.1.2.2) If the verification fails *and* it's not the final consonantal cluster of the word (string of graphemes), check if the cluster in question can be divided into an initial and a final cluster according to the Onset First Principle and verify these as legal ones.

1.2) Strip any unstressed prefixes found and mark stressed ones (loop).

1.2.1) **stripObetPref()**. Latin and Greek prefixes that are not stressed (Thorell 1981 p 60). Strip these.

1.2.2) If there are no more unstressed prefixes to strip, check if there is any stressed prefix to mark.

1.2.2.1) **markBetPref()**. Verb particles and certain other prefixes carry stress. If any of these is found, mark it with a '/' before the stressed vowel.

1.3) If no stressed prefix is to be found, over to suffixes.

1.3.1) **markBetSuf()**. Try to find a suffix that carries stress, if so mark it.
These are Greek suffixes and some other suffixes.

1.3.2) Strip gender, definiteness and plural one at a time and after each check if there is any suffix that carries stress, 1.3.2.2, to mark (loop).

1.3.2.1) **stripObetSuf()**. Strip first gender, then definiteness and last plural. In Swedish these suffixes always appear in this order and only, at most, one of each.

1.3.2.2) **markBetSuf()**. See 1.3.1.

1.3.3) **stripObetSuf()**. If any suffix that carries stress can be found, strip it. This time the Rhyme-matcher looks for noun, adjective, verb and adverb derivational suffixes as well as noun, adjective and verb inflectional endings.

1.3.4) **markBetSuf()**. See 1.3.1. This time, for the last time.

1.4) If a stress has been placed in what the Rhyme-matcher believes to be the stem (i.e. what is left after the stripping process), paste the stress marker ("/") into the string of graphemes we started with, in the same position as in the stem, and go to 1.6.

1.5) If no stress has been placed, try to place one, using the Onset First Principle, looking for a heavy syllable or gemination.

1.5.1) **markeraStavelse()**. Go through each and every consonant cluster in the string (loop). If there is only one vowel, go to 1.5.1.4.

1.5.1.1) Grab medial or final cluster and check for gemination, if found, go to 1.5.1.5 and mark this syllable for stress (see section 3.4).

1.5.1.2) If no gemination, use the Onset First Principle to divide the cluster, if medial, into an initial and a final cluster and verify these. If a valid final cluster is identified, go to 1.5.1.5 and mark this syllable for stress.

1.5.1.3) If not medial, verify final cluster. If valid, go to 1.5.1.5 and mark this syllable for stress.

1.5.1.4) If only one vowel is found it naturally gets the stress.

1.5.1.5) If a position for the stress has been found, mark that position. If not, mark the first syllable in the string.

1.5.2) See step 1.4.

1.5.3) If no stress has been placed yet, the word isn't adhering to Swedish phonotactics. This shouldn't be able to happen though since a test for this already has been performed in step 1.1 (efficiency reasons).

1.6) Return the analysed string of graphemes.

5 Results

Unfortunately, we didn't have the time to test the Rhyme-matcher the way we would have wanted. Every little step on the way took much longer than we ever could have imagined. All from coming to an agreement on the topic of this paper, background research and getting settled at ELD to moulding the SUC, *Stockholm Umeå Corpus*, into a workable format, programming the actual Rhyme-matcher and programming around Lingo-specific limitations, was eating more and more time from this paper. Also our ambition took us further than time permitted. However, it is our strong belief that this is good grounds to build on. These matters will be discussed in sections 6 and 7.

Below are a couple of actual outputs from the Rhyme-matcher showing it in action. First a session running on a portion of the SUC occurrence list containing words with the initial letter |r|.

```
-- "Startord: sm/art"
-- "rahle kan jag inte säga!"
-- "rahlfs kan jag inte säga!"
-- "rahmbeck kan jag inte säga!"
-- "ralph kan jag inte säga!"
-- "ramirez kan jag inte säga!"
-- "ramqvists kan jag inte säga!"
-- "Rimord: r/art"
-- "rawitz kan jag inte säga!"
-- "Rimord: realiserb/art"
-- "rechtlich kan jag inte säga!"
-- "rechtslehre kan jag inte säga!"
-- "rechtswissenschaft kan jag inte säga!"
-- "refresh kan jag inte säga!"
```

(Output from the Rhyme-matcher)

As we can see above the Rhyme-matcher finds two matching rhymes, **r/art** and **realiserb/art**. We also see a list of words, highlighted in italic, that the Rhyme-matcher deems unpronounceable in Swedish. One of the problems with the Rhyme-matcher is that it doesn't show what it lets get through. It could, of course, but the output would be too extensive to be to any use. You would need to use search and reformatting tools which we, at least at the moment, do not have. During the development phase such tools are essential to establish if the rules are too lax as well as too tense.

```
-- "Startord: sn/urvlar"
-- "F:  I: rvl"
-- "F: r I: vl"
-- "F: rv I: l"
-- "Rimord: m/urvlar"

-- "Startord: pl/ugga"
-- "F:  I: tst"
-- "F: t I: st"
-- "Rimord: hetst/ugga"

-- "Startord: str/essa"
-- "Rimord: prins/essa"

-- "Startord: finem/ang"
-- "Rimord: lavem/ang"
```

(Selected output from the Rhyme-matcher)

Above we first have an example of the Onset First Principle function in action. The **F:** represents the current final cluster and the **I:** the initial. This kind of output can be switched on and of but to have it on while scanning large lists of words for a match would also, sad enough, generate a too extensive output to be useful without special tools. Next we see an example of when the Onset First Principle goes wrong. Above we also see two examples of matching word-to-word. Note that these outputs are somewhat edited.

6 Discussion

The purpose of the Rhyme-matcher was to construct a program that could both entertain and educate children. No children have yet been allowed to play with the program and it hasn't been incorporated into any of ELD's titles. None the less, we believe that with additional work on the program (see below) it could be a complement in an educational CD-ROM title. So far the program does what it is supposed to do, it matches rhymes. It does not match false rhymes, it only permits strings of characters that adhere to Swedish phonotax and it always produces some kind of output. We believe that the idea of matching of non-words with Swedish lexem's could entertain children, at least if they have some reading and writing abilities. However, if the program ever were to be tested with children they would most certainly bring additional flaws of the machine to our attention. They would, of course, be the best judges of the usability of the machine.

We must also analyse if we could have achieved our purpose in any other form or way. For example, instead of listing all consonantal clusters in a database, we could have worked from a rule-based system. Rules that govern what characters can precede or follow any chosen character. What we would have gained would certainly have been scientific elegance and perhaps clarity of structures. We don't know if a rule-based system would have been more effective but we can agree on that it would have been more tasteful.

In the making of the Rhyme-matcher several questions and problems arose. A number of problems ascending from the decision to leave out any phonetical or phonological representation had to be considered. How should we represent the phonemes, that is, which graphemes corresponded to which phonemes? What number of permitted members in final consonantal sequences should we allow? Since we did not incorporate any database of permitted medial sequences, how should we handle problems arising from deciding whether a medial sequence was to be permitted or not? How should we master the power of the Rhyme-matcher, what affixes should we use, how should we list them and in what order should the machine strip them?

6.1 Orthographical representation

The problems surrounding the orthographical representation have surely meant that the database is incomplete. A number of missing grapheme clusters was discovered in the testing of the Rhyme-matcher. Genuine Swedish lexical items weren't permitted and the sole reason was that either final or initial consonantal clusters within the word weren't represented in the database. We naturally added them to the database. One of the missing initial clusters were |hj| that phonetically is represented by /j/ and we simply missed the |h|. Further testing might reveal additional missing clusters. There were not only drawbacks to the choice of not incorporating a phonological description. We, as already mentioned, gained time and didn't have to make any of the hard decisions regarding choice of phoneme. Speakers of a language sometimes differ quite dramatically in their pronunciation depending on dialect, sociolect etc. and that makes an exhaustive phonological description almost impossible.

6.2 Affixes and consonantal clusters

The number of permitted members in final consonantal clusters in the database is restricted to 4. Therefore longer sequences are disallowed by the machine even though they exists in Swedish. The choice to restrict the number of permitted members was obvious. Longer sequences are secondary and scarce and would increase our database with a great number of additional consonantal clusters. The performance of the machine would probably not be affected by the inclusion of these clusters.

As been represented in sections 3.2 and 3.3.1 there are affixes that sometimes carry stress and sometimes don't. To solve this problem we had to make a choice of either always assigning stress to these ambivalent affixes or never. We choose to never assign stress to them, they are simply stripped whenever found.

The way of listing the affixes also had to be considered. We choose to order them by length and status of their members e.g. we put the derivational suffix *-iell*, before *–ell*, otherwise the machine would never find *–iell*, because obviously its three last members would already have been stripped and only *–i* would remain. This problem also arose with the cluster database and we solved it in a similar manner. We put the longest sequences first in the database, thereby not risking to chop genuine long sequences into shorter ones e.g. |skl| before |sk|.

6.3 The Onset First Principle and language specific reasons

There is, as one may have noticed when reading 4.1, some problems with the Onset First Principle. One of these problems is the fact that the maximum allowed onset, in Swedish, is not always the onset that is intuitively perceived. Take, for example, the case of *hörslarna*, 'hearings'. In this case the Onset First Principle, and the Rhyme-matcher, deems the correct segmentation to be |hör-slarna| while intuitively the segmentation should be |hörs-larna|. This because the Rhyme-matcher first tries |hö-rslarna| and discovering that |rsl| is not an allowed initial consonant cluster in Swedish, continues with |hör-slarna|. Here it is encountered with the fact that |sl| is an allowed initial cluster and stops there, taking no concern about morphological or semantical aspects of syllable segmentation. Another good example is *byggnadsarbetsnämndemännens,* '-', even though it is a compund and not a simple word. Below is an output from the Rhyme-matcher showing how it handles the clusters |ggn|, in the context |by-ggn-ad|, and |tsn|, in the context |arbe-tsn-ämnd|.

```
-- "F:  I: ggn"
-- "F: g I: gn"
-- "F:  I: tsn"
-- "F: t I: sn"
-- "Startord: byggnadsarbetsn/ämndemännens"
```

(Selected output from the Rhyme-matcher)

6.4 Efficiency

It can be discussed whether it is more efficient to run through a heap of medial clusters (each and every combination of initial and final clusters) instead of the current implementation. The problem is that this would not only mean a very long list of (medial) clusters to run through, but it would also demand special treatment of the final cluster. It can also be discussed if it would be more efficient to skip the cluster test, as it is now, on the medial and final clusters and jump directly to the segmentation according to the Onset First Principle and test for both initial and final clusters instead. We do not believe this to be the case since the segmentation of each cluster takes time, and have to run through the clusters anyway. If we can get away without using the Onset First Principle we save time.

7 Further improvements

To further improve the Rhyme-matcher it will have to be more extensively tested as is. Other methods for testing will also have to be implemented. One improvement of the machine would be to incorporate the much-debated phonological description. This would make it possible to give the users of the Rhyme-matcher access to a phonematical description of every input that hopefully would give valuable clues to graphemic and phonemic correspondence. This might be useful in the child's process of learning to master the written language. Another improvement is to equip the Rhyme-matcher with additional speech synthesis. To give every input to the machine a sound would vastly improve its educational potential. The speech synthesis would add auditory feedback to every input and the combination of visual and auditive feedback would surely further improve the benefits of the machine.

There are some language-specific reasons that we believe would improve the Rhyme-matcher if incorporated. For a discussion concerning these, see section 4.1. For example could the problem cluster |ggn|, in the context |by-ggn-ad|, in *byggnadsarbetsnämndemännens* (section 6.1) be handled by introducing a rule that handles gemination. In the case of the cluster |tsn|, in the context |arbe-tsn-ämnd|, the problem lies in handling the presence of a suture-s (fog-s) which only appears in compounds. We have only had the time to implement and test the use of allowed, initial and medial/final, (consonant) clusters as language-specific reasons (see sections 3.2.1 and 5) to set aside the Onset First Principle but we are well aware of that a more intuitively sound syllable segmentation could probably be yielded by introducing more rules of exception.

We would also like to incorporate some kind of Part Of Speech analysis since a word's Part Of Speech category in some cases affects the stress pattern.

Some sort of lexicon containing semantic information, for handling exceptions rules can't handle, would also be nice to include. This to handle difficulties with so called minimal pairs (see section 6.2). Along the way one might find more exceptions needed to be handled by the lexicon, for example corporate and personal names etc.

References

Bruce, Gösta, 1998. *Allmän och svensk prosodi*. Lund: Reprocentralen, Lunds universitet.

Dahlerus, Lars, 1994. *Om språkpauser och stavelseindelning*. Lidingö: Bokförlaget Instructor.

Elert, Claes-Christian, 1964. *Phonologic studies of quantity in Swedish: based on material from Stockholm speakers*. Uppsala [Stockholm]: Språkförlaget Skriptor.

Håkansson, Marie, Stenquist, Annika, 1989. *Om uttal*. Kungälv: Skriptor Förlag.

Katamba, Francis, 1989. *An introduction to Phonology*. Longman Group UK Limited.

Lindau, Mona. 1970. *Prosodic problems in a generative phonology of Swedish. Working Papers 2, Phonetic Laboratory*. Lund: Lund University.

Malmberg, Bertil, 1956. *Svensk fonetik: i jämförande framställning.* Lund: Gleerups Förlag.

Noreen, Adolf, 1907-1910. *Vårt språk: nysvensk grammatik i utförlig framställning, andra bandet*. Lund: Gleerups Förlag.

Oldberg, Ragnar, 1945. *En bok om rim*. Lund: C.W.K. Gleerups Förlag.

Sigurd, Bengt, 1965. *Phonotactic structures in Swedish*. Lund: Berlingska boktryckeriet.

Spencer, Andrew, 1996. *Phonology*. Padstow, Cornwall: Blackwell Publishers Ltd.

Söderbergh, Ragnhild, 1985. *Barnets tidiga språkutveckling*. Stockholm: Liber förlag.

Thorell, Olof, 1981. *Svensk ordbildningslära*. Stockholm: Nordstedts tryckerier.

Trask, Robert, Lawrence, 1993. *A dictionary of gramatical terms in linguistics*. Padstow, Cornwall: Blackwell Publishers Ltd.

E-mail, Levande Böcker, Gunnel Källgren, 98-03-09