



Language Manual

Swedish

Ann-Marie and Ingmar

Language Manual
Swedish
Ann-Marie and Ingmar
Edition 40
1 January 2005

This document was produced by Acapela Group. We welcome and consider all comments and suggestions. Please send them to:

Acapela Group
Box 1328
SE-171 26 Solna
Sweden

Phone +46 (0) 8 799 86 00
Fax + 46 (0) 8 799 86 01

Acapela Group
33, Boulevard Dolez
7000 Mons
Belgium

Tel: +32 (0)65 37 42 75
Fax: +32 (0)65 37 42 76

Acapela Group
3939, la Lauragaise
BP 58309
F-31683 Labège cedex
France

Tel: +33 (0)5 62 24 71 00
Fax: +33 (0)5 62 24 71 01

www.acapela-group.com

© Copyright Acapela Group 2005. All rights reserved.

List of contents

| | | |
|-------|--|----|
| 1 | General | 5 |
| 1.1 | Notational conventions | 5 |
| 2 | Letters in orthographic text | 6 |
| 2.1 | Characters treated as letters in other languages | 6 |
| 3 | Non-alphanumeric characters | 7 |
| 3.1 | Punctuation characters | 7 |
| 3.1.1 | Comma, colon and semicolon | 7 |
| 3.1.2 | Quotation marks | 7 |
| 3.1.3 | Full stop | 7 |
| 3.1.4 | Question mark | 8 |
| 3.1.5 | Exclamation mark | 8 |
| 3.1.6 | Parentheses | 8 |
| 3.2 | Characters whose pronunciation varies | 8 |
| 3.2.1 | Hyphen | 8 |
| 3.2.2 | Equals sign | 9 |
| 3.2.3 | Asterisk | 9 |
| 3.2.4 | Multiple occurrences of the same character | 9 |
| 3.3 | Other non-alphanumeric characters | 10 |
| 3.4 | Control characters | 10 |
| 3.5 | Characters ignored by the system | 11 |
| 4 | Number processing | 12 |
| 4.1 | Full number pronunciation | 12 |
| 4.2 | Leading zero | 12 |
| 4.3 | Grouping of numbers | 13 |
| 4.4 | Decimal numbers | 13 |
| 4.5 | Currency amounts | 13 |
| 4.6 | Time of day | 14 |
| 4.7 | Year reading | 14 |
| 4.8 | Dates | 15 |
| 4.9 | Arithmetic operators | 15 |
| 4.10 | Ordinal numbers | 15 |
| 4.11 | Mixed digits and letters | 16 |
| 5 | Swedish Phonetic Text | 17 |
| 5.1 | Consonants | 18 |
| 5.2 | Comments on phonetic symbols for consonants | 18 |
| 5.2.1 | Retroflex sounds | 18 |
| 5.2.2 | Aspiration | 19 |
| 5.3 | Vowels | 20 |
| 5.4 | Comments on phonetic symbols for vowels | 20 |
| 5.4.1 | Ä and Ö in front of R-sounds | 20 |
| 5.5 | Extra symbols for phonetic details | 20 |
| 5.5.1 | Lexical stress | 21 |
| 5.5.2 | Emphasis and reduction | 21 |
| 5.5.3 | Foreign sounds | 21 |
| 5.5.4 | Punctuation marks | 22 |
| 5.5.5 | Hyphen | 22 |
| 5.5.6 | Underscore | 22 |
| 6 | The RULSYS phonetic alphabet | 23 |
| 6.1 | RULSYS Consonants | 24 |
| 6.2 | Comments on phonetic symbols for consonants | 24 |
| 6.2.1 | Retroflex sounds | 24 |
| 6.2.2 | Aspiration | 25 |
| 6.3 | RULSYS Vowels | 26 |
| 6.4 | Comments on phonetic symbols for vowels | 26 |
| 6.4.1 | Long and short vowels | 26 |
| 6.4.2 | Ä and Ö in front of R-sounds | 27 |
| 6.5 | Extra symbols for phonetic details | 27 |
| 6.5.1 | Lexical stress | 27 |
| 6.5.2 | Emphasis and reduction | 27 |
| 6.5.3 | Foreign sounds | 27 |

| | |
|--|----|
| 6.5.4 Hyphen | 28 |
| 6.5.5 Space..... | 28 |
| 7 How to change pronunciation errors..... | 29 |
| 7.1 Change the orthography | 29 |
| 7.1.1 Spelling incorrectly..... | 29 |
| 7.1.2 Use of hyphen | 29 |
| 7.1.3 Expanding acronyms..... | 29 |
| 7.2 Using phonetic text..... | 29 |
| 7.2.1 Choosing the right phonetic symbols..... | 29 |
| 8 Abbreviations..... | 30 |
| 8.1 Context sensitive abbreviations..... | 31 |

1 General

This document discusses certain aspects of text-to-speech processing for the Swedish text-to-speech system, in particular the different types of input characters and text that are allowed.

This version of the document corresponds to the High Density voices Ann-Marie and Ingmar.

1.1 Notational conventions

The following notational conventions are used in this manual:

- For linguistic entities in general, **boldface** is used.
- Input text is written in a non proportional font.
- Output text is written in *italics*.
- Keyboard entities are written within angle brackets < >.
- Phonetic transcriptions are written within slashes or hash marks depending on the phonetic alphabet used.

The following abbreviations are used in this manual:

| | |
|----|---------------|
| LM | Letter mode |
| SM | Sentence mode |

See the User's Guide for a description of the two different reading modes. Note that Sentence mode sometimes is referred to as Normal mode.

2 Letters in orthographic text

Characters from A- and a-o may constitute a word. Certain other characters are also considered as letters, notably those used as letters in other European languages, i.e. “n, o, c, é”, see section 2.2.

Characters outside of these ranges, i.e. digits and non-alphanumeric characters such as punctuation characters and currency markers etc, are not considered as letters. If such a non-letter is included within a word, the word is ended where the non-letter appears and the following letters considered belonging to a new word.

2.1 Characters treated as letters in other languages

Characters with the accent marks < ´ ` ^ > are considered to be letters. In LM, the name of the accent mark is pronounced. In SM, the accent marks affect the pronunciation in some cases, see examples below.

| Expression | Reading (LM) | Reading (SM) |
|------------|----------------------------------|--------------------------------|
| ó | <i>o med akut accent</i> | <i>o med akut accent</i> |
| à | <i>a med grav accent</i> | <i>a med grav accent</i> |
| î | <i>i med cirkumflex accent</i> | <i>i med cirkumflex accent</i> |
| moské | <i>m o s k e med akut accent</i> | <i>moské</i> |

Other characters used as letters in other European languages are pronounced according to Table 1.

| Letter | LM | SM (next to other letters) | SM (in isolation) |
|--------|--------------------------|-------------------------------|----------------------|
| Ç | <i>(stort) c cedille</i> | <i>s</i> | <i>c cedille</i> |
| ü | <i>tyskt y</i> | <i>y</i> | <i>y</i> |
| ç | <i>c cedille</i> | <i>s</i> | <i>s</i> |
| ë | <i>e trema</i> | <i>e</i> | <i>e</i> |
| ï | <i>i trema</i> | <i>i</i> | <i>i</i> |
| ÿ | <i>y trema</i> | <i>y</i> | <i>y</i> |
| Ü | <i>(stort) tyskt y</i> | <i>y</i> | <i>y</i> |
| ñ | <i>n tilde</i> | <i>nj</i> | <i>n tilde</i> |
| Ñ | <i>(stort) n tilde</i> | <i>nj</i> | <i>n tilde</i> |
| ß | <i>beta</i> | <i>s</i> | <i>beta</i> |

Table 1 Characters used as letters in other European languages

3 Non-alphanumeric characters

The processing of non-alphanumeric characters varies, depending on the reading mode, context of the character, and its function within that context. These are the types of non-alphanumeric characters to be distinguished:

- Characters always processed as punctuation, and having a direct effect on the intonation and pausing in SM.
- Non-punctuation characters that are always pronounced, with no effect on the intonation or pausing.
- Characters whose pronunciation varies according to context.

Below is a discussion of the characters grouped by type. For each character, the pronunciation is given in the three basic reading modes.

3.1 Punctuation characters

Table 2 below lists punctuation characters permitted in the normal text input string and their readings in LM. In SM the punctuation marks are silent but they affect both rhythm and intonation as described in the sections below.

| Character | LM |
|-----------|--------------------------------|
| , | <i>komma</i> |
| : | <i>kolon</i> |
| ; | <i>semikolon</i> |
| “ | <i>citationstecken</i> |
| « | <i>vänster citationstecken</i> |
| » | <i>höger citationstecken</i> |
| . | <i>punkt</i> |
| ? | <i>frågetecken</i> |
| ! | <i>utropstecken</i> |
| (| <i>vänsterparentes</i> |
|) | <i>högerparentes</i> |

Table 2 Punctuation characters

3.1.1 Comma, colon and semicolon

Comma < , >, colon < : > and semicolon < ; > cause a brief pause where they occur in a sentence. Comma is used as decimal marker, see section 4.4. Colon is used in currency amounts and in time expressions, see section 4.5 and 4.6, and in ordinals, see section 4.10.

3.1.2 Quotation marks

Quotes < “ ” > or < « » > appearing around a single word or a group of words cause a brief pause before and after the quoted text.

3.1.3 Full stop

A full stop < . > is a sentence terminal punctuation mark which causes a falling end-of-sentence intonation pattern and is accompanied by a somewhat longer pause. A full stop may also be used as a decimal marker in a number (see section 4.4).

3.1.4 Question mark

A question mark <?> ends a sentence and causes a major fall in intonation followed by a pause.

3.1.5 Exclamation mark

The exclamation mark <!> has the same effect as the full stop, causing a major fall in intonation followed by a pause.

3.1.6 Parentheses

Parenthesis < () > around a single word or a group of words cause a brief pause before and after the bracketed text.

3.2 Characters whose pronunciation varies

The pronunciation of the characters listed below varies according to their context.

| Character | LM | SM |
|-----------|----------------------|-------------------------------------|
| - | <i>streck</i> | (se 3.2.1 och 3.2.4) |
| = | <i>likhetstecken</i> | (se 3.2.2 och 3.2.4) |
| + | <i>plus</i> | <i>plus</i> (se även 3.2.4) |
| * | <i>stjärna</i> | (se 3.2.3 och 3.2.4) |
| # | <i>nummertecken</i> | <i>nummertecken</i> (se även 3.2.4) |

Table 3 Characters with varying pronunciation

All examples below show the reading in SM.

3.2.1 Hyphen

The reading of hyphen <-> follows the following principles:

- If a digit immediately follows, it is pronounced *minus*.
- When part of some of the date formats, it is pronounced *till*, see also section 4.7.
- If surrounded by spaces or digits, it is pronounced as a short pause.
- When used to mark compound words, it is not pronounced.
- When preceded by a letter and followed by a letter, digit, or punctuation mark, it is not pronounced.
- Hyphen is discarded at the end of a line, and causes the two parts of the hyphenated word to be joined into a single word.
- In most other cases, hyphen is pronounced *streck*.

Expression

44-3
15-20 oktober
6-10 nov
åren 1998-2004

550522 - 4758
2000-07-31
Det menar du inte - är det sant?
data-
maskin
data-
En mening kan vara ovan-
ligt lång
--bc

Reading

44 minus 3
femtonde till tjugonde oktober
sjätte till tionde november
åren nittonhundra nittioåtta till
tjugohundra fyra
550522 (paus) 4758.
tjugohundra (paus) noll sju (paus) trettioett
det menar du inte (paus) är det sant?
datamaskin*

data
en mening kan vara ovanligt lång.

streck streck B C

*In LM, there will be a pause between the words, i.e. *data* (paus) *maskin*.

3.2.2 Equals sign

Equals sign $< = >$ is always pronounced *lika med* if preceded and followed by digits. In all other cases, it is pronounced *likhetstecken*.

Expression

$2 * 3 = 6$
 $c b = b c$

Reading

två gånger tre är lika med sex
C B likhetstecken B C

3.2.3 Asterisk

Asterisk $< * >$ is pronounced *gånger* if a digit immediately precedes and follows it; it is pronounced *stjärna* in all other cases.

Expression

$2 * 3$
 $* b c$

Reading

två gånger tre
stjärna B C

3.2.4 Multiple occurrences of the same character

In SM, if more than three of the same character occur in sequence without a space separating the characters, only the first three occurrences will be pronounced. This is only valid for the following characters: $< * + - = \# >$.

Expression

+++++

=====

#####

Reading

stjärna stjärna stjärna
plus plus plus
streck streck streck
likhetstecken likhetstecken likhetstecken
nummertecken nummertecken nummertecken

3.3 Other non-alphanumeric characters

Each of the characters in Table 4 are read in LM, most of them also in SM.

| Character | LM | SM |
|-----------|---|---|
| % | <i>Procent</i> | <i>procent</i> |
| & | <i>Och</i> | <i>och</i> |
| / | <i>snedstreck</i> | <i>snedstreck</i> |
| ' | <i>Apostrof</i> | <i>apostrof</i> |
| @ | <i>snabel-a</i> | <i>snabel-a</i> |
| < | <i>Vänsterhake</i> | <i>vänsterhake</i> |
| > | <i>Högerhake</i> | <i>högerhake</i> |
| _ | <i>Understryk</i> | <i>understryk</i> |
| ^ | <i>pil upp</i> | <i>pil upp</i> |
| # | <i>nummertecken</i> | <i>nummertecken</i> |
| \$ | <i>dollar</i> | <i>dollar</i> |
| £ | <i>Pund</i> | <i>pund</i> |
| € | <i>euro</i> | <i>euro</i> |
| ¿ | <i>upp-och-nervänt frågetecken</i> | <i>upp-och-nervänt frågetecken</i> |
| ¡ | <i>upp-och-nervänt utropstecken</i> | <i>upp-och-nervänt utropstecken</i> |
| ½ | <i>en halv</i> | <i>en halv</i> |
| ¼ | <i>en fjärdedel</i> | <i>en fjärdedel</i> |
| ~ | <i>Tilde</i> | <i>tilde</i> |
| ± | <i>plus minus</i> | <i>plus minus</i> |
| ² | <i>Kvadrat</i> | <i>kvadrat (see the special case below and section 8.1)</i> |
| | <i>Blanktecken</i> | <i>(tyst)</i> |

Table 4 Other non-alphanumeric characters

Special case

km², mm², cm² and m² are read as *kvadratkilometer, kvadratmillimeter* etc. (**km**, and **cm** are read as abbreviations, see chapter 8, but **m** on its own is read as the name of the letter; to have it read as e.g. *Meter* it is necessary to enter it in the user lexicon).

3.4 Control characters

In LM (only) the control characters <RETURN> and <TAB> among others are read out in all systems according to Table 5 below. In most cases, they are read as *kontroll* + the appropriate letter, e.g. ^T is read as *kontroll T*. Some other control characters are read according to their function in text applications, see below for examples.

| Character | LM |
|-------------------|---------------------|
| ^H <BACKSPACE> | <i>radera</i> |
| ^I <TAB> | <i>tab</i> |
| ^J <LINE FEED> | <i>ny rad</i> |
| ^K <VERTICAL TAB> | <i>vertikal tab</i> |
| ^M <RETURN> | <i>retur</i> |
| ^[<ESC> | <i>escape</i> |

Table 5 Control characters

3.5 Characters ignored by the system

All characters that are not described in chapter 2 and 3 and that are not phonetic symbols or digits, are ignored by the system. Normally, these characters are omitted but some of them may cause the sentence they appear in to be silent.

4 Number processing

Strings of digits that are sent to the text-to-speech converter are processed in several different ways, depending on the reading mode, format of the digit string, and the immediately surrounding punctuation or non-numeric characters. To familiarise the user with the various types of formatted and non-formatted strings of digits that are recognized by the system, we provide below a brief description of the basic number processing along with examples.

Number processing is subdivided into the following categories:

- 4.1 Full number pronunciation
- 4.2 Leading zero
- 4.3 Grouping of numbers
- 4.4 Decimal numbers
- 4.5 Currency amounts
- 4.6 Time of day
- 4.7 Year reading
- 4.8 Dates
- 4.9 Arithmetic operators
- 4.10 Ordinal numbers
- 4.11 Mixed digits and letters

The examples in this chapter show the reading in SM. In LM, all digits and punctuation marks are read.

4.1 Full number pronunciation

Full number pronunciation is given for the whole number part of the digit string, i.e. the part to the left of the decimal marker.

Numbers denoting thousands, millions and billions (numbers larger than 999) may be grouped using space or full stop. The highest number read is 999999999999 (twelve digits). Numbers higher than this are read as separate digits, with pauses between groups.

| Number | Reading |
|--------------|--|
| 2425 | <i>tvåtusenyrahundratjugofem</i> |
| 1000000000 | <i>en miljard</i> |
| 123456789012 | <i>ettundratjugotremiljarder fyrahundrafemtiosexmiljoner sjuhundraåttioniotusen tolv</i> |
| 2.425 | <i>tvåtusenyrahundratjugofem</i> |
| 22 000 | <i>tjugotvåusen</i> |

4.2 Leading zero

Numbers that begin with 0 (zero) are read digit by digit, with pauses between groups of digits if there are five or more digits. Numbers beginning with zero followed by exactly two digits where the first is not a zero, the last two digits are read as a full number.

| Number | Reading |
|--------|------------------------------------|
| 09253 | <i>noll nio två (paus) fem tre</i> |
| 020 | <i>noll tjugo</i> |

4.3 Grouping of numbers

Numbers denoting thousands, millions etc (numbers larger than 999) may be grouped using space or full stop. In order to achieve the right pronunciation the grouping must be done in the following way:

- Numbers are grouped in groups of three starting at the end.
- The first group in a number may consist of one, two, or three digits.
- If a group, other than the first, does not contain exactly three digits, the sequence of digits is not interpreted as a full number.

| Number | Reading |
|------------|---|
| 12.345 | <i>tolvtusentrehundrafyrtiofem</i> |
| 12.34.5 | <i>tolv punkt trettiofyra punkt fem</i> |
| 12.345.234 | <i>tolvmiljoner trehundrafyrtiofemtusen tvåhundra trettiofyra</i> |
| 12 345 234 | <i>tolvmiljoner trehundrafyrtiofemtusen tvåhundra trettiofyra</i> |

4.4 Decimal numbers

Full stop or comma (colon in currency amounts) are recognised as decimal markers in the Swedish system. If full stop is used, the decimal part is read according to the number processing rules described above. Note that a number containing full stop is not read as a decimal number if the full stop is followed by exactly three digits. If comma is used and the following digits are more than two, the decimal part will be read as separate digits.

| Number | Reading |
|---------|---|
| 16,234 | <i>sexton komma två tre fyra</i> |
| 3,1415 | <i>tre komma ett fyra ett fem</i> |
| 1251,04 | <i>ettusent två hundrafemtioett komma noll fyra</i> |
| 2,50 | <i>två komma femtio</i> |
| 2.50 | <i>två punkt femtio</i> |
| 3.141 | <i>tretusen etthundrafyrtioett</i> |

4.5 Currency amounts

The following principles are valid for currency amounts in the Swedish system:

- Numbers with two decimal places preceded or followed by the currency marker kr is read as monetary amounts
- Numbers with two decimal places preceded by the currency markers £, \$, or € are read as monetary amounts.
- A space is optional between amount and currency marker.
- The accepted decimal markers are colon < : >, comma < , > and full stop < . >.
- The decimal part (consisting of two digits) in monetary amounts is read as “*och nn öre/pence/cents*” respectively.
- If the decimal part is “00” it will be read as “*och noll öre/pence/cents*” respectively.
- The sequence < :- > is not read.

| Expression | Reading |
|------------|--|
| 2988:45 | <i>två tusen niohundra åttioåtta och fyrtiofem</i> |
| 200:- | <i>två hundra</i> |
| 21:50 | <i>tjugoen och femtio</i> |
| 0:45 | <i>noll och fyrtiofem</i> |
| 1:99 | <i>en och nittionio</i> |

| | |
|---------------|---------------------------------------|
| kr 20:50 | <i>tjugo kronor och femtio öre</i> |
| kr 200:00 | <i>tvåhundra kronor och noll öre</i> |
| kr 200:- | <i>tvåhundra kronor</i> |
| 200Kr | <i>tvåhundra kronor</i> |
| 200:00Kr | <i>tvåhundra kronor och noll öre</i> |
| 200:00 kronor | <i>tvåhundra och noll noll kronor</i> |
| 200Kr50 | <i>tvåhundra kronor femtio</i> |
| £20:50 | <i>tjugo pund och femtio pence</i> |
| \$20:00 | <i>tjugo dollar och noll cent</i> |
| 20:00\$ | <i>tjugo dollar</i> |
| 20:00 dollar | <i>tjugo dollar</i> |
| €20:50 | <i>tjugo euro och femtio cent</i> |
| €20:00 | <i>tjugo euro och noll cent</i> |

4.6 Time of day

Time of day is read in SM if the following format is observed:

kl H:M:S

H represents the hours, **M** represents the minutes, **S** represents the seconds. The seconds are optional. **kl** is pronounced *klockan*. Hours, minutes and seconds must be separated by colon < : > or full stop < . > which both are pronounced *och*. The *och* before minutes is not read if the minutes are equal to **00**, nor is it read if the seconds are equal to **00**.

| Expression | Reading |
|-------------------|--|
| kl 20:50 | <i>klockan tjugo och femtio</i> |
| kl 21:00 | <i>klockan tjuoett noll noll</i> |
| kl 21.45.00 | <i>klockan tjuoett fyrtiofem och noll noll</i> |
| kl 21.00.45 | <i>klockan tjuoett noll noll och fyrtiofem</i> |

4.7 Year reading

A four digit number between 1100 and 2099 is read as hundreds (year reading) in the following cases:

- The number is preceded by one of the words **ar**, **aret**, **aren**. If such expression is preceded by < - > and two or four digits, the expression is recognised as a year interval. No spaces are allowed around < - >.
- The number is preceded by a month name.
- The number is preceded by a date followed by a month name or an abbreviation for a month.
- The number is included in one of the following formats: 2003/10/03, 2003-10-03, 3/10 2003, see section 4.8.

In all other cases, numbers between 2000 and 2099 are read as thousands. Numbers between 1100 and 1999 are normally read as years, except when the numbers contain decimals.

| Expression | Reading |
|-------------------|--|
| året 1988 | <i>året nittonhundraåttioåtta</i> |
| åren 1939-45 | <i>åren nittonhundratrettionio till fyrtiofem</i> |
| åren 1998-2010 | <i>åren nittonhundranittioåtta till tjugohundratio</i> |
| år 2000 | <i>år tjugohundra</i> |
| år 2004 | <i>år tjugohundra fyra</i> |
| 1088 | <i>ettusen åttioåtta</i> |
| 1900 | <i>nittonhundra</i> |
| 1988 | <i>nittonhundraåttioåtta</i> |
| 2000 | <i>tvåusen</i> |
| 1988,0 | <i>ettusen niohundraåttioåtta komma noll</i> |
| 1988.032 | <i>ettusen niohundraåttioåtta punkt noll trettio två</i> |
| 13 sep 2004 | <i>trettonde september tjugohundrafyra</i> |

4.8 Dates

A sequence of digits and hyphens or slashes is read as a date if one of the formats below is followed:

yyyy-mm-dd
yyyy/mm/dd
dd/mm yyyy

yyyy must be a four digit number representing a year between **1800** and **2099**.

mm must be a month number between **01** and **12**.

dd must be a day number between **01** and **31**.

Expression

1898-10-13
1977-03-09
2045-05-16
1999/10/30
3/4 2009

Reading

artonhundra nittioåtta tio tretton
nittonhundra sju noll tre noll nio
tjugohundra fyrtio fem noll fem sexton
nittonhundra nittionio tio trettio
tredje i fjärde tjugohundra nio

4.9 Arithmetic operators

Numbers together with arithmetical operators are read according to the examples below. See also 3.2.

Expression

-12
+24
2*3
25%
3,4%
10.30

Reading

minus tolv
plus tjugofyra
två gånger tre
tjugofem procent
tre komma fyra procent
tio punkt trettio

4.10 Ordinal numbers

Numbers are read as ordinals in the following cases:

- the number is followed by a month name or an abbreviation for a month and is lower than or equal to the allowed number of days in the particular month.
- the number is followed by colon < : > and < e > (< a > after 1 and 2).

Ordinal numbers are recognised in SM only.

Expression

31 januari
2 feb
32 januari

Reading

trettioförsta januari
andra februari
trettio två januari

Note that the system is not able to detect where a sequence of digits (for example a phone number) ends, and where a date begins. However, the problem does not occur in the following cases:

- the sequence of digits contains more than two digits without any spaces in between.
- a punctuation mark, for example a comma, is placed between (e.g.) the phone number and the date.
- the number preceding the date is larger than the number of allowed days in the month.

Also note that the system is unable to recognise spaces, tabs etc between colon and letters. This means the sequence **2:<tab>a** is read *andra*. See examples on next page.

Expression

ring 764 35 02 feb

ring 7643502 feb

1:e

2:a

2: a

3:e

3:a

Reading*ring sjuhundrasextiofyra trettiofem andra februari**ring sju miljoner sexhundrafyrtiotretusen femhundra två feb**förste**andra**andra**tredje**trea***4.11 Mixed digits and letters**

If a letter appears within a sequence of digits, the groups of digits will be read as numbers according to the rules above. The letter marks the boundary between the numbers. The letter will also be read.

Expression

77B84Z3

0092B87-B

Reading*sjuttiosju B åttiofyra Z tre**noll noll nio två B åttiosju B*

5 Swedish Phonetic Text

In the current version of the text-to-speech system, SAMPA (Speech Assessment Methods Phonetic Alphabet) is used when making lexicons or using phonetic strings within texts. In earlier versions, RULSYS was used. For the voices based on RULSYS, a conversion is made automatically from SAMPA to RULSYS inside the system.

We recommend new users to use only SAMPA since this is the notation that will be used in future development. Users who are already familiar with the RULSYS alphabet still have the possibility to use it when making user lexicons for all RULSYS-based voices (among them the Swedish voices Ann-Marie and Ingmar). There will be a description of RULSYS in the next chapter.

For the sake of clarity, SAMPA transcriptions are written within slashes (/ /) and RULSYS transcriptions within hash marks (# #). Note that neither the slashes nor the hash marks are part of the actual transcription.

The Swedish system uses a phonetic alphabet similar to the Swedish subset of SAMPA. The phonetic alphabet is described below.

If the pronunciation is incorrect the user may write phonetic transcriptions in the text. Then, a PRN-tag is needed to switch to phonetic mode, see User's Guide. It is also possible to make user lexicons (see User's Guide), or change the orthography of a word (see chapter 7) in order to achieve the preferred pronunciation.

5.1 Consonants

Table 6 lists the phonetic symbols used for the Swedish consonants along with example words (the letters corresponding to the consonant sound are in boldface) and their transcriptions.

| Consonant symbol | Example | Transcription |
|------------------|-------------------|-------------------------|
| b | bar | /b A:4 r/ |
| d | dag | /d A:4 g/ |
| g | gul | /g } :4 l/ |
| p | spar, lapp | /s p A:4 r/, /l a4 p/ |
| t | stal, hatt | /s t A:4 l/, /h a4 t/ |
| k | skal, tack | /s k A:4 l/, /t_h a4 k/ |
| l | lus | /l } :4 s/ |
| m | mor | /m u:4 r/ |
| n | natt | /n a4 t/ |
| N | ung | /u4 N/ |
| f | far | /f A:4 r/ |
| v | var | /v A:4 r/ |
| s | sand | /s a4 n d/ |
| S | skjul | /S } :4 l/ |
| C | kjol | /C u:4 l/ |
| j | jul | /j } :4 l/ |
| h | hatt | /h a4 t/ |
| r | ris | /r i:4 s/ |
| rs | kors | /k O4 rs/ |
| rt | hårt | /h o:4 rt/ |
| rl | sorl | /s o:4 rl/ |
| rn | barn | /b A:4 rn/ |
| rd | bord | /b u:4 rd/ |

Table 6 Swedish consonant symbols in SAMPA

5.2 Comments on phonetic symbols for consonants

5.2.1 Retroflex sounds

The pronunciation of the dental consonants **T**, **D**, **S**, **L** and **N** is changed when preceded by **R**. The resulting sounds are called ‘supradentals’ or ‘retroflex’. Compare the pronunciation of **bord** and **bod**, **fort** and **fo!** This pronunciation of **T**, **D**, **S**, **L**, and **N** is represented by the merging of /r/ and the following consonant symbol, i.e. /r t/ becomes /rt/.

| | | |
|----------|-------------|------------|
| Examples | hart | /h o:4 rt/ |
| | bord | /b u:4 rd/ |
| | kors | /k O4 rs/ |
| | sorl | /s o:4 rl/ |
| | barn | /b A:4 rn/ |

Note that **R** also may affect more than one of the following consonants.

| | | |
|----------|-----------------|-------------------------|
| Examples | bordsduk | /b u:3 rd rs rd } :1 k/ |
|----------|-----------------|-------------------------|

5.2.2 Aspiration

In Swedish, the voiceless stop sounds **P, T, K** are typically *aspirated* in certain positions of the word. That is, they are followed by a “puff of breath”.

Aspiration occurs in the following contexts:

- Initially in all words if followed by **a)** a vowel or **b)** one of the consonants **J, L, N, R, V** followed by a vowel.
- In all grave and acute non-compound words in front of a primary stressed vowel unless **P, T, K** are preceded by **S**. One of the consonants **J, L, N, R, V** is allowed between the aspirated stop and the vowel.
- Morpheme initially in compounds followed by a vowel. One of the consonants **J, L, N, R, V** are allowed between the aspirated stop and the vowel. The previous morpheme may end with an **S**.
- Within a morpheme in a compound followed by **a)** a primary stressed vowel or **b)** one of the consonants **J, L, N, R, V** followed by a primary stressed vowel. The aspiration does not occur if **P, T, K** are preceded by **S**.

Aspiration does not occur in the following contexts:

- After **S** within the same syllable.
- In front of /@/
- In grave non-compound words in front of vowel with secondary stress.

The synthesis cannot always correctly predict when a voiceless stop is to be aspirated. The user can indicate the aspiration of a voiceless stop by placing /_h/ after the stop symbol.

Note that prefixes like be-, for-, an- etc do not affect the aspiration.

| | | | | |
|-----------|--------------------|---------------------------------|--------------|----------------|
| Examples: | Aspiration: | | | |
| | pall | /p_h a4 l/ | plats | /p_h l a4 t s/ |
| | katt | /k_h a4 t/ | prov | /p_h r U4 v/ |
| | tall | /t_h a4 l/ | tvang | /t_h v O4 N/ |
| | potatis | /p_h U t_h A:4 t I s/ | | |
| | traplatta | /t_h r E:3 p_h l a1 t a/ | | |
| | skalpotatis | /s k A:3 l p_h U t_h A:1 t I s/ | | |
| | tegeltak | /t_h e:3 g @ l t_h A:1 k/ | | |
| | hustak | /h } : 3 s t_h A:1 k/ | | |
| | anta | /a3 n t_h A:1/ | | |
| | No aspiration: | | | |
| | spar | /s p A:4 r/ | | |
| | skatt | /s k a4 t/ | | |
| | stall | /s t a4 l/ | | |
| | hoppa | /h O3 p a1/ | | |
| | hoppat | /h O3 p a1 t/ | | |
| | ljusstake | /j } :3 s s t A:1 k @/ | | |

Note that the retroflex /rt/ may be aspirated.

Example: **parti** /p_h a rt_h i:4/

5.3 Vowels

Table 7 lists the phonetic symbols used for the Swedish vowels along with example words and their transcriptions.

| Vowel symbol | Example | Transcription |
|--------------|---------|---------------|
| A: | lam | /l A:4 m/ |
| a | lamm | /l a4 m/ |
| e: | se | /s e:4/ |
| e | sett | /s e4 t/ |
| @ | natten | /n a4 t @ n/ |
| i: | dit | /d i:4 t/ |
| I | ditt | /d I4 t/ |
| u: | bo | /b u:4/ |
| U | bott | /b U4 t/ |
| }: | hus | /h }:4 s/ |
| u | hund | /h u4 n d/ |
| y: | sy | /s y:4/ |
| Y | sytt | /s Y4 t/ |
| o: | gå | /g o:4/ |
| O | gått | /g O4 t/ |
| E: | häl | /h E: l/ |
| E | häll | /h E4 l/ |
| {: | här | /h {:4 r/ |
| { | herr | /h {4 r/ |
| 2: | lös | /l 2:4 s/ |
| 2 | lös | /l 24 s/ |
| 9: | för | /f 9:4 r/ |
| 9 | för | /f 94 r/ |

Table 7 Swedish vowel symbols in SAMPA

5.4 Comments on phonetic symbols for vowels

5.4.1 Ä and Ö in front of R-sounds

A and Ö have special phonetic symbols when preceding an R. These symbols are /{: / and /9: / respectively for long vowels and /{/ / and /9/ / respectively for short vowels.

| | | |
|----------|------|-----------|
| Examples | har | /h {:4 r/ |
| | herr | /h {4 r/ |
| | for | /f 9: r/ |
| | forr | /f 94 r/ |

5.5 Extra symbols for phonetic details

In the current version of the Swedish synthesis certain phonetic details can be specified in phonetic text. This can be exploited in case the user wishes to achieve an unusual pronunciation, or if the transcription automatically generated by the system is inaccurate.

5.5.1 Lexical stress

In words with more than one syllable, one (and normally only one) of the syllables is more prominent than the others. This is referred to as word stress, or lexical stress. Words of one syllable also have word stress when spoken in isolation, although many may lose the stress in certain contexts.

In Swedish, lexical stress is often used to distinguish between two words that are spelled alike, but pronounced differently. It is therefore very important to include stress marks when writing transcriptions.

There are two word accents in Swedish: acute and grave. The acute accent is in SAMPA denoted by a < 4 > placed immediately after the stressed vowel. The grave accent is denoted by a < 3 > placed immediately after the stressed vowel (with primary stress) and a < 1 > immediately after the vowel that receives secondary stress. All compounds have grave accent. Examples:

| | | | |
|----------|------------------------------|-------------------------|----------------|
| Examples | anden (the bird) | /a4 n d @ n/ | (acute accent) |
| | anden (in the bottle) | /a3 n d @1 n/ | (grave accent) |
| | andpastej | /a3 n d p_h a1 s t e j/ | (compound) |

5.5.2 Emphasis and reduction

It is also possible to emphasise or reduce the stress on a particular word in a phrase or sentence. In the input text string, this is done by placing < _X >, where X represents a single digit between 0 and 9, within a PRN-tag (see User's Guide) immediately before the word whose prominence is to be altered. The emphasis mark can also be used in transcriptions in a user lexicon.

| | |
|---------|--|
| _2 | normal stress for most words |
| _0 | makes a word non-stressed |
| _1 | gives stress to a normally unstressed word |
| _3 - _9 | gives levels of emphatic stress |

5.5.3 Foreign sounds

A few non-Swedish sounds, which sometimes occur in the pronunciation of foreign words and names, are permitted in the transcriptions of the current version of the Swedish synthesis. Depending on which voice is chosen, these sounds will be rendered more or less similar to the foreign speech sounds. The sounds in question are exemplified in Table 8.

| Phonetic symbol | Example | Transcription | Language of origin |
|-----------------|---------------------|------------------------------|--------------------|
| _ch t S | Churchill | /_ch t S 2:4 _ch t S I l/ | English |
| _dh D | this | /_dh D I4 s/ | English |
| _er 9: | circle | /s _er 9:4 k @ l/ | English |
| _jh d S | John | /_jh d S O4 n/ | English |
| _ow O v | Windows | /_w v I4 n d _ow O v _Z s/ | English |
| _th T | Thatcher | /_th T {4 _ch t S @ r/ | English |
| _W v | Windows | /_w v I4 n d _ow O v _Z s/ | English |
| _Z s | Windows | /_w v I4 n d _ow O v _Z s/ | English |
| _zh S | Jeanne d'Arc | /_zh S A:4 n _ d a4 r k/ | French |
| _ng N | Mont Blanc | /m O4 _ng N _ b l A:4 _ng N/ | French |
| _aa A: | Maas | /m _aa A:4 s/ | German |
| _x k | Bach | /b a4 _x k/ | German |
| | Loch | /l O4 _x k/ | Scottish English |

Table 8 Foreign sounds in SAMPA

5.5.4 Punctuation marks

The punctuation marks < . ! ? , > used in phonetic text have the same effect on intonation as when appearing in orthographic text. In SAMPA the punctuation marks are denoted /_/, /_!_/, /_? _/, and /_com _/ respectively.

5.5.5 Hyphen

In phonetic text, hyphen (in SAMPA underscore + hyphen, < _- >) can be used to separate parts of a compound word. If the hyphen separating two parts of a word comes at the end of a line, the word is not read until the second part on the next line is typed. For a description of the use of the hyphen character in normal orthographic text, see section 3.3.1.

5.5.6 Underscore

Underscore, < _ > or < _spc > may be used to improve the pronunciation of phonetic transcriptions. It is useful when transcribing names and abbreviations for example.

Examples: **Jeanne d’Arc** /_zh S A:4 n _ d a4 r k/
 oae / u: _spc A: _spc e:4/

6 The RULSYS phonetic alphabet

Note that we recommend new users to use only SAMPA since this is the notation that will be used in future development. Note also that it is only possible to use RULSYS when making user lexicons, not in the input text string.

The following differentiates RULSYS from SAMPA in the Swedish system:

- no spaces are used in transcriptions
- the lexical accent is placed before the vowel to be stressed, not after as in SAMPA
- there are symbols for blocking retroflex sounds
- there are a number of symbols for foreign sounds

Note that the hash marks (# #) are used to indicate RULSYS transcriptions and to differentiate them from SAMPA transcriptions; the hash marks are not part of the actual transcriptions.

If the pronunciation is incorrect the user may write phonetic transcriptions in the text. Then, a PRN-tag is needed to switch to phonetic mode, see User's Guide. It is also possible to make user lexicons (see User's Guide), or change the orthography of a word (see chapter 7) in order to achieve the preferred pronunciation.

6.1 RULSYS Consonants

Table 8 lists the phonetic symbols in RULSYS used for the Swedish consonants along with example words and their transcriptions.

| Consonant symbol | Example | Transcription |
|------------------|-------------------|------------------|
| B | bar | #B'A:R# |
| D | dag | #D'A:G# |
| G | gul | #G'U:L# |
| P | spar, lapp | #SP'A:R#, #L'AP# |
| T | stal, hatt | #ST'A:L#, #H'AT# |
| K | skal, tack | #SK'A:L#, #T'AK# |
| L | lus | #L'U:S# |
| M | mor | #M'O:R# |
| N | natt | #N'AT# |
| NG | ung | #'UNG# |
| F | far | #F'A:R# |
| V | var | #V'A:R# |
| S | sand | #S'AND# |
| SJ | skjul | #SJ'U:L# |
| TJ | kjol | #TJ'O:L# |
| J | jul | #J'U:L# |
| H | hatt | #H'AT# |
| R | ris | #R'I:S# |
| 2S | kors | #K'Å2S# |
| 2T | hårt | #H'Å:2T# |
| 2L | sorl | #S'Å:2L# |
| 2N | barn | #B'A:2N# |
| 2D | bord | #B'O:2D# |

Table 8 RULSYS consonants

6.2 Comments on phonetic symbols for consonants

6.2.1 Retroflex sounds

For a description of retroflex sounds, see 5.2.1. In RULSYS, aspiration is denoted by < 2 > plus the consonant in question. Note that the R is left out, see examples below.

| | | |
|----------|-------------|----------|
| Examples | hart | #H'Å:2T# |
| | bord | #B'O:2D# |
| | kors | #K'Å2S# |
| | sorl | #S'Å:2L# |
| | barn | #B'A:2N# |

Note that R also may affect more than one of the following consonants.

| | | |
|----------|-----------------|-------------------|
| Examples | bordsduk | #B''O:2D2S-D'U:K# |
|----------|-----------------|-------------------|

In most cases, R + dental consonant is automatically converted by the system to the corresponding retroflex sound. If one wish to keep R + dental consonant as two separate sounds, there is the possibility to use #rx# instead of #R#.

| | | |
|---------|-----------------|-----------------|
| Example | barrskog | #B''Arx-SK'O:G# |
|---------|-----------------|-----------------|

6.2.2 Aspiration

Aspiration is described in section 5.2.2. In RULSYS, aspiration is denoted by adding an **h** to the phonetic symbol for the stop to be aspirated.

Example **hustak** #H"U:STh'A:K# (#T# is aspirated and denoted #Th#)

6.3 RULSYS Vowels

Table 9 lists the phonetic symbols in RULSYS used for the Swedish vowels along with example words and their transcriptions.

| Vowel symbols | Example | Transcription |
|---------------|---------|---------------|
| Å: | lam | #L'Å:M# |
| Å | lamm | #L'AM# |
| E: | se | #S'E:# |
| E | sett | #S'ET# |
| E0 | natten | #N'ATEON# |
| I: | dit | #D'I:T# |
| I | ditt | #D'IT# |
| O: | bo | #B'O:# |
| O | bott | #B'OT# |
| U: | hus | #H'U:S# |
| U | hund | #H'UND# |
| Y: | sy | #S'Y:# |
| Y | sytt | #S'YT# |
| Å: or]: | gå | #G'Å:# |
| Å or] | gått | #G'ÅT# |
| Ä: or [: | häl | #H'Ä:L# |
| Ä or [| häll | #H'ÄL# |
| Ä3 or [3 | här | #H'Ä3R# |
| Ä4 or [4 | herr | #H'Ä4R# |
| Ö: or \: | lös | #L'Ö:S# |
| Ö or \ | löss | #L'ÖS# |
| Ö3 or \3 | för | #F'Ö3R# |
| Ö4 or \4 | förr | #F'Ö4R# |

Table 9 RULSYS vowels

6.4 Comments on phonetic symbols for vowels

6.4.1 Long and short vowels

In order to differentiate between long and short vowels, a colon < : > is used to denote long vowels.

| | | |
|---------|-------------|---------|
| Example | tak | #T'Å:K# |
| | tack | #T'AK# |

6.4.2 Ä and Ö in front of R-sounds

A and have special phonetic symbols when preceding an R. These symbols are #Ä3# and #Ö3# respectively for long vowels and #Ä4# and #Ö4# respectively for short vowels.

| | | |
|----------|-------------|---------|
| Examples | har | #H'Ä3R# |
| | herr | #H'Ä4R# |
| | for | #F'Ö3R# |
| | forr | #F'Ö4R# |

6.5 Extra symbols for phonetic details

In the current version of the Swedish synthesis certain phonetic details can be specified in phonetic text. This can be exploited in case the user wishes to achieve an unusual pronunciation, or if the transcription automatically generated by the system is inaccurate.

6.5.1 Lexical stress

Lexical stress is described in 5.5.1. In RULSYS, acute accent is denoted by < ' > and grave accent by < “ >. Both are placed immediately in front of the vowel with primary stress. Compounds have < “ > in front of the vowel with primary stress and < ‘ > in front of the vowel with secondary stress.

| | | | |
|----------|------------------------------|----------------|----------------|
| Examples | anden (the bird) | #'ANDEON# | (acute accent) |
| | anden (in the bottle) | #"ANDEON# | (grave accent) |
| | andpastej | #"AND-PAST'EJ# | (compound) |

For the use of hyphen in compounds, see section 6.5.4.

6.5.2 Emphasis and reduction

In RULSYS, phrase level stress is denoted in the same way as in SAMPA. See section 5.5.2.

6.5.3 Foreign sounds

A few non-Swedish sounds, which sometimes occur in the pronunciation of foreign words and names, are permitted in the transcriptions of the current version of the Swedish synthesis. Depending on which voice is chosen, these sounds will be rendered more or less similar to the foreign speech sounds. The sounds in question are exemplified in Table 10.

| Phonetic symbol | Example | Transcription | Language of origin |
|-----------------|---------------------|-----------------|--------------------|
| ch | Churchill | #ch'Ö:chIL# | English |
| dh | this | #dh'IS# | English |
| er | circle | #S'erKE0L# | English |
| jh | John | #jh'ÄN# | English |
| ow | Windows | #W'INDowZ# | English |
| th | Thatcher | #th'Ä4chE0R # | English |
| W | Windows | #W'INDowZ# | English |
| Z | Windows | #W'INDowZ# | English |
| zh | Jeanne d'Arc | #zh'A:N D'ARK# | French |
| ng | Mont Blanc | #M'Äng BL'A:ng# | French |
| aa | Maas | #M'aaS# | German |
| X | Bach | #B'AX# | German |
| | Loch | #L'ÄX# | Scottish English |

Table 10 Symbols for foreign sounds

6.5.4 Hyphen

The hyphen < - > is used in transcriptions to separate parts of a compound word in order to achieve the proper pronunciation. The stressed vowel in the last part of the compound should be preceded by an apostrophe < ' >.

Example **symaskin** #S"Y:-MASJ' I:N#

It can also be used to avoid the misinterpreting of consonant sounds. Since there are no spaces in RULSYS transcriptions, the sequences #NG#, #SJ#, and #TJ# are ambiguous since these as well as #N#, #G#, #S#, #J# and #T# are phonetic symbols. A hyphen or an asterisk can be used to separate the symbols.

Exemple **bangard** #B"A:N-G' Å:2D#
 vagnshjul #V"ANGNS-J'U:L#
 nattjobb #N"AT-J' ÅB#

6.5.5 Space

Space < > or full stop < . > may be used to improve the pronunciation of phonetic transcriptions. It is useful when transcribing names and abbreviations for example.

Examples: **Jeanne d'Arc** #zh'A:N D'ARK#
 oae #O: A: 'E:#

7 How to change pronunciation errors

Words that are not pronounced correctly by the text-to-speech converter can be entered in the user lexicon (see User's guide). There are two ways to do this: either, the user enters a phonetic transcription of the word (see chapter 6), or, the user rewrites the word orthographically. Phonetic transcriptions can also be entered directly in the text, using a PRN-tag (see User's guide).

7.1 Change the orthography

7.1.1 Spelling incorrectly

It is possible to intentionally misspell a word by trying to spell a word in a more phonetic manner, i.e., choosing non-ambiguous letter combinations to represent difficult sounds. A long vowel can be made short by doubling the following consonant.

| | | | |
|----------|---------------|-------------------|---------------|
| Examples | extern | can be misspelled | extarn |
| | juice | | jös |
| | juste | | sjyst |
| | pub | | pubb |

7.1.2 Use of hyphen

A hyphen character can be used within a word to separate two letters that might otherwise be incorrectly pronounced together.

| | | | |
|---------|----------------|----------------|-----------------|
| Example | isjakt | can be written | is-jakt |
| | krakagg | | krak-agg |

7.1.3 Expanding acronyms

Not very many acronyms are handled by the current Swedish system (see chapter 8). Therefore, it may be very useful to expand them in the user dictionary. Since acronyms should be expanded to more than one word it may be difficult to enter a proper transcription. It is much easier to enter the words in question orthographically.

| | | |
|----------|-----------|-------------------|
| Examples | DN | Dagens Nyheter |
| | SJ | Statens Järnvagar |

7.2 Using phonetic text

When you are unable to correct a pronunciation error by misspelling the word, phonetic text should be used to produce the desired pronunciation. When phonetic text is used, the system bypasses the normal spelling pronunciation rules, and pronounces each phonetic symbol "literally", according to the examples listed in Tables 6 and 7.

7.2.1 Choosing the right phonetic symbols

A helpful way to transcribe in phonetic text is to work with a dictionary. Normally, dictionaries give the pronunciation for each word. They also provide a pronunciation key to show how to pronounce the special symbols used in the pronunciation guide. Similarly, Tables 6 and 7 give the pronunciation key for the special phonetic symbols used in Swedish for the text-to-speech converter.

Using a dictionary, look up the word you want to transcribe. Next to the word you should find the pronunciation. Working with the dictionary's pronunciation key and Tables 6 and 7, convert the dictionary pronunciation symbols to the appropriate Swedish symbols for the text-to-speech converter. Symbols that are used in the dictionary to mark syllable or word boundaries should be ignored. Be sure to include the stress assignment information since lexical stress is an important part of a word's pronunciation.

8 Abbreviations

In the current version of the Swedish text-to-speech system, the following abbreviations are recognized in all contexts (in SM only). These abbreviations are case-sensitive and require no full stop in order to be recognised as abbreviations. If a full stop accompanies the abbreviation, the sentence is terminated at the abbreviation. There are also some context sensitive abbreviations, see section 8.1.

The user lexicon may be used to redefine any of these abbreviations, or to create your own.

| Abbreviation | LM | SM |
|--------------|----------------|---------------------------------|
| allm | <i>A L L M</i> | <i>Allmän</i> |
| amp | <i>A M P</i> | <i>Ampere</i> |
| ang | <i>A N G</i> | <i>Angående</i> |
| ankn | <i>A N K N</i> | <i>Anknytning</i> |
| anm | <i>A N M</i> | <i>Anmärkning</i> |
| avd | <i>A V D</i> | <i>Avdelning</i> |
| bilj | <i>B I L J</i> | <i>Biljett</i> |
| bitr | <i>B I T R</i> | <i>Biträdande</i> |
| bost | <i>B O S T</i> | <i>Bostad</i> |
| ca | <i>C A</i> | <i>Cirka</i> |
| cm | <i>C M</i> | <i>Centimeter (se även 3.3)</i> |
| co | <i>C O</i> | <i>Company</i> |
| div | <i>D I V</i> | <i>Diverse</i> |
| dr | <i>D R</i> | <i>Doktor</i> |
| dvs | <i>D V S</i> | <i>det vill säga</i> |
| dyl | <i>D Y L</i> | <i>Dylikt</i> |
| eftr | <i>E F T R</i> | <i>Efterträdare</i> |
| enl | <i>E N L</i> | <i>Enligt</i> |
| etc | <i>E T C</i> | <i>et cetera</i> |
| ev | <i>E V</i> | <i>Eventuellt</i> |
| exkl | <i>E X K L</i> | <i>Exklusive</i> |
| exp | <i>E X P</i> | <i>Expedition</i> |
| fr | <i>F R</i> | <i>från</i> |
| hkr | <i>H K R</i> | <i>Hästkrafter</i> |
| hr | <i>H R</i> | <i>Herr</i> |
| inkl | <i>I N K L</i> | <i>Inklusive</i> |
| inst | <i>I N S T</i> | <i>Institution</i> |
| jr | <i>J R</i> | <i>Junior</i> |
| kbm | <i>K B M</i> | <i>Kubikmeter</i> |
| kg | <i>K G</i> | <i>Kilo</i> |
| kl | <i>K L</i> | <i>Klockan</i> |
| km | <i>K M</i> | <i>Kilometer (se även 3.3)</i> |
| kr | <i>K R</i> | <i>Kronor</i> |
| kvm | <i>K V M</i> | <i>Kvadratmeter</i> |
| mg | <i>M G</i> | <i>Milligram</i> |
| milj | <i>M I L J</i> | <i>Miljon</i> |
| mkr | <i>M K R</i> | <i>Miljoner kronor</i> |
| mr | <i>M R</i> | <i>Mister</i> |
| mrs | <i>M R S</i> | <i>Missis</i> |
| nr | <i>N R</i> | <i>Nummer</i> |
| omkr | <i>O M K R</i> | <i>omkring</i> |
| osv | <i>O S V</i> | <i>och så vidare</i> |
| pga | <i>P G A</i> | <i>på grund av</i> |
| prel | <i>P R E L</i> | <i>Preliminär</i> |
| proc | <i>P R O C</i> | <i>Procent</i> |
| resp | <i>R E S P</i> | <i>Respektive</i> |
| sekr | <i>S E K R</i> | <i>Sekreterare</i> |
| stud | <i>S T U D</i> | <i>Studerande</i> |
| tekn | <i>T E K N</i> | <i>Teknisk</i> |

Table 11 Recognised abbreviations in the Swedish system

8.1 Context sensitive abbreviations

The following abbreviations are only recognised if containing spaces and sentence mode is enabled: **t ex** (till exempel) and **bl a** (bland annat). The abbreviation **pga** can also be written **p g a** but then it is only recognised in sentence mode.

Abbreviations for days of the week and month are only recognised in sentence mode and no full stop is allowed (except when in sentence final position). Days of the week abbreviations are only expanded when preceding a date followed by a month or a month abbreviation. Month abbreviations are expanded in the following contexts:

- after a date (20 feb)
- after a date interval (10-20 apr)

Accepted abbreviations for days of the week are:

man, tis, ons, tor, fre, lor, son.

Accepted abbreviations for months are:

jan, feb, febr, mar, apr, jun, jul, aug, sep, sept, okt, nov, dec.

Abbreviations for days and months are only expanded if the date is in accordance with the number of days allowed in the month in question.

Some abbreviations denoting measure and space units (**ml, cl, dl, mm** and **dm**) are only expanded when preceded by digits. In all other cases, **mm** is read *medmera*. Telephone related abbreviations (**tel, tfn, mob**) are only expanded when there are digits following, no full stops are allowed. Both types of abbreviations are only recognised in sentence mode.

| Expression | Reading |
|------------|----------------------------------|
| t ex | <i>till exempel</i> |
| bl a | <i>bland annat</i> |
| tis 13 okt | <i>tisdag trettonde oktober</i> |
| 13 okt | <i>trettonde oktober</i> |
| 13-14 okt | <i>tretton minus fjorton okt</i> |
| 10 ml | <i>tio milliliter</i> |

m is normally read as the name of the letter. By making a user dictionary and adding some predefined transcriptions (see Table 12 on the following page), it is possible to have **m** read in the following four ways:

- **m** is read *meter* only when preceded by digits
- **m** is read *meter* only in the case **m²**
- **m** is always read *meter*
- **m** is always read as the name of the letter

| Dictionary entry | Transcription in RULSYS & SAMPA | Example | Reading (SM only) |
|------------------|---------------------------------|-------------------|--------------------------|
| M | #@GM#, /_@ g m/ | 12 m | <i>tolv meter</i> |
| | | 12 m ² | <i>tolv kvadratmeter</i> |
| | | m ² | <i>Kvadratmeter</i> |
| | | l m n | <i>l m n</i> |
| M | #@QM#, /_@ q m/ | 12 m | <i>tolv m</i> |
| | | 12 m ² | <i>tolv kvadratmeter</i> |
| | | m ² | <i>Kvadratmeter</i> |
| | | l m n | <i>l m n</i> |
| M | #@ZM#, /_@ z m/ | 12 m | <i>tolv meter</i> |
| | | 12 m ² | <i>tolv kvadratmeter</i> |
| | | m ² | <i>Kvadratmeter</i> |
| | | l m n | <i>l meter n</i> |
| M | #@M#, /_@ m/ | 12 m | <i>tolv m</i> |
| | | 12 m ² | <i>tolv m kvadrat</i> |
| | | m ² | <i>m kvadrat</i> |
| | | l m n | <i>l m n</i> |

Table 12 Special transcriptions for M