Language Manual

HQ and CO North Sami

Language Manual: HQ and CO North Sami

Published 10 February 2015

Copyright © 2014-2015 Acapela Group.

This document was produced by Acapela Group. We welcome and consider all comments and suggestions. Please, use the *Contact Us* link at our website: http://www.acapela-group.com

Table of Contents

1 General

This document discusses certain aspects of text-to-speech processing for the North Sami text-to-speech system, in particular the different types of input characters and text that are allowed.
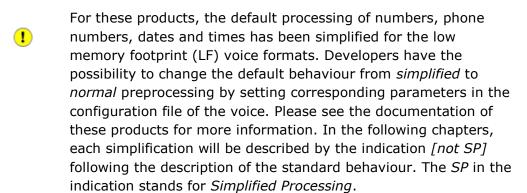
This version of the document corresponds to the High Quality (HQ) and Colibri (CO) North Sami voices.

Please note that the *User's Guide*, mentioned several times in the manual, is called *Help* in some applications.

Note: This language manual is general and applies to all Acapela Group North Sami voices specified above. One or more of the voices may be included in a certain Acapela Group product.

> Note: For efficiency reasons, the processing described in this document has a different behaviour in some Acapela Group products. Those products are:
>
> - Acapela TTS for Windows Mobile
> - Acapela TTS for Linux Embedded
> - Acapela TTS for iOS
> - Acapela TTS for Android
>
> For these products, the default processing of numbers, phone numbers, dates and times has been simplified for the low memory footprint (LF) voice formats. Developers have the possibility to change the default behaviour from *simplified* to *normal* preprocessing by setting corresponding parameters in the configuration file of the voice. Please see the documentation of these products for more information. In the following chapters, each simplification will be described by the indication *[not SP]* following the description of the standard behaviour. The *SP* in the indication stands for *Simplified Processing*.

2 Letters in orthographic text

Characters from *A-ZÁÅÄÆÖØČŦÐŠŽŊ* and *a-záåäæöøčŧđšžŋ* may constitute a word. Certain other characters are also considered as letters, notably those used as letters in other European languages, e.g. *é, è, ô, ü*. These letters are not pronounced as in their native languages though, they are pronounced as regular *e, o, u* etc. However, when one of these letters stands on its own, they are read with an indication of the diacritic (modification of the letter), for instance (for the letters mentioned above): *e akuhttaakseanttain*, *e gravisakseanttain*, *o sirkumfleaksaakseanttain*, *u čuoggáiguin*.

Characters outside of these ranges, i.e. numbers, punctuation characters and other non-alphanumeric characters, are not considered as letters.

3 Punctuation characters

Punctuation marks appearing in a text affect both rhythm and intonation of a sentence. The following punctuation characters are permitted in the normal input text string: *, : ; " ". ? ! ( ) { } [ ]*

3.1 Comma, colon and semicolon

Comma ',', colon ':' and semicolon ';' cause a brief pause to occur in a sentence, accompanied by a small rising intonation pattern just prior to the character.

3.2 Quotation marks

Quotes ' " " ' appearing around a single word or a group of words cause a brief pause before and after the quoted text.

3.3 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 chapter *Number processing*) and in abbreviations (see chapter *Abbreviations*).

3.4 Question mark

A question mark '*?*' ends a sentence and causes question-intonation, first rising and then falling.

3.5 Exclamation mark

The exclamation mark '*!*' is treated in a similar manner to the full stop, causing a falling intonation pattern followed by a pause.

3.6 Parentheses, brackets and braces

Parenthesis '*( )*', brackets '*[ ]*' and braces '*{ }*' appearing around a single word or a group of words cause a brief pause before and after the bracketed text.

4 Other non alphanumeric characters

4.1 Non-punctuation characters

The characters listed below are processed as non-letter, non-punctuation characters. Some are pronounced at all times and others are only pronounced in certain contexts, which are described in the following sections of this chapter.

Non-punctuation characters

Symbol	Reading
/	finjusáhcu
+	plussa
$	dollár
£	punda
€	euro
¥	yen
<	unnit go
>	stuorit go
%	proseanta
^	sirkumfleaksa
|	ceakko njulgessáhcu
~	tilde
@	smarfe a
^{2}	see below
^{3}	see below
*	see below
-	see below
=	see below

4.2 The ^{2} and ^{3} signs

The reading of expressions with ^{2} and ^{3} is:

Expression	Reading
mm^2	kvadráhttamillimehter
cm^2	kvadráhttasentimehter
m^2	kvadráhttamehter
km^2	kvadráhttakilomehter
mm^3	guđahasmillimehter
cm^3	guđahascentimehter

Expression	Reading
m³	guđahasmehter
km³	guđahaskilomehter

4.3 Symbols whose pronunciation varies depending on the context

4.3.1 Hyphen

A hyphen '-' is pronounced *minus* in two cases:

1. if followed by a digit and no other digit is found in front of the hyphen, i.e. as in the pattern -X but not in X-Y or X -Z where X, Y, and Z are numbers.
2. if followed by a digit and an equals sign '=', i.e. as in the pattern X-Y=Z. Space is allowed between digits, hyphen and equals sign.

If there is no equals sign, as in X-Y or X -Z, the hyphen is pronounced *sáhcu*.

In certain date formats, it is pronounced *gitta*. If more than one hyphen appear together only one is processed.

Expression	Reading	
-3	minus golbma	
44-3	njealljeloginjeallje sáhcu golbma	[Not SP]
44-3=41	njealljeloginjeallje minus golbma lea njealljelogiokta	[Not SP]
44 - 3 = 41	njealljeloginjeallje minus golbma lea njealljelogiokta	
skábmam. 15. b. - njukčam. 15. b.	skábmamánu vihttanuppelogát beaivvis njukčamánu vihttanuppelogát beaivái	[Not SP]
jagit 1998-2004	jagit ovccinuppelotčuođiovccilogigávcci gitta guokteduhátnjeallje	[Not SP]
vuorká-dávvirat	vuorká-dávvirat	
31-07-2008	suoidnemánu golbmalogivuosttaš beaivi guokteduhátgávcci	

4.3.2 Asterisk

Asterisk '*' is pronounced *geardde* if enclosed by digits and followed by equals sign '='. In other cases it is pronounced *asterisk*.

Expression	Reading
2*3	guokte asterisk golbma
2*3=6	juokte geardde golbma lea guhtta
*bc	asterisk b c

4.3.3 Equals sign

Equals sign '=' is pronounced *lea* if followed by a digit. In all other cases it is pronounced *ovttamađodatmearka*.

Expression	Reading
2*3=6	guokte geardde golbma lea guhtta
cd=dc	c d ovttamađodatmearka d c

5   Number processing

The text-to-speech system applies the so-called new counting principles for Sami, which means that, for instance, 27 is read as *guoktelogičieža* and not as *čiežagoalmmatlohkái*.

Strings of digits that are sent to the text-to-speech converter are processed in several different ways, depending on the format of the string of digits 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 recognised 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:

The inflection of numbers and abbreviations
Full number pronunciation
Leading zero
Decimal numbers
Currency amounts
Ordinal numbers and date formats
Arithmetic operators
Mixed digits and letters
Time of day
Years
Phone numbers

5.1 The Inflection of numbers and abbreviations

The North Sami text-to-speech system includes a facility for inflecting numbers and abbreviations linked to numbers and currencies. The case ending must be separated from the number or the abbreviation by a colon. The system inflects the whole combination according to the case indicated by the ending. The system supports cardinal numbers in all singular and plural cases. However, for *:a* that can be both accusative and genitive the latter is used. Ordinal numbers are only handled in date formats, see *Ordinal numbers and date formats*. [not SP]

The ending recognised for the various cases are: [not SP]

Case	Ending	Example	Reading
Singular:			
Nominative		6	guhtta
Accusative	:a	6:a	guđa
Genitive	:a	6:a	guđa
Illative	:i	6:i	guhtti
Locative	:s	6:s	guđas
Comitative	:in	6:in	guđain
Essive	:n	6:n	guhttan
Plural:			
Nominative	:t	6:t	guđat
Accusative	:id	6:id	guđaid
Genitive	:id	6:id	guđaid
Illative	:ide	6:ide	guđaide
Locative	:in	6:in	guđain
Comitative	:iguin	6:iguin	guđaiguin
Essive	:n	6:n	guhttan

Abbreviations handled properly when followed by case ending and preceded by a number:

$ £ € ¥ ¢

USD GBP E e NOK SEK DKK MK mk FIM fim KR kr RU ru

dollár punda pund euro yen márki

mm cm dm m. km ml cl dl l. hl mg cg dg g. hg kg

ha kj j mj kcal w mw kw kwh

m2 m² m3 m³ cm2 cm² cm3 cm³ mm2 mm² mm3 mm³

dm2 dm² dm3 dm³ km2 km² km3 km³

§ % °c sek ms

Expression	Reading
6 cm:i	guhtta centimehterii
10 $:a	logi dollára
20 punda:s	guoktelogi pundas
23 dm:in	guoktelogigolmmain desimehteriin
40 kw:n	njealljelohkin kilowáttan

An extended set of case endings, some more informal, is recognized after regular words and abbreviations. The ending is read as being part of the word and the word/abbreviation itself is not inflected.

:a	:aid	:ain	:an	:as	:at	:din	:dis
:eš	:i	:id	:idda	:ide	:iguin	:ii	:iidda
:iin	:in	:is	:n	:s	:sán	:žan	:sis
:t	:ái	:áin	:árat	:áriid	:š	:ža	:žii
:žiid	:žiidda	:žiiguin	:žiin	:žis			

That is, "New York:s" is not read "New Yorkas" but "New Yorks" and "Norge:in" is not read "Norgiin" but "Norgein".

5.2 Full number pronunciation

Full number pronunciation is given for the whole number part of the digit string.

Example

2425	full number
2.425	full number
2 425	full number
24,25	24 is a full number, 25 is the decimal part

Numbers denoting thousands, millions and billions (numbers larger than 999) may be grouped using space or full stop (not comma). In order to achieve the right pronunciation the grouping must be done correctly.

The rules for grouping of numbers are the following:

- 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.
- The highest number read is 999999999999 (twelve digits). Numbers higher than this are read as separate digits.
- An exception is made for year pronunciation, which occurs in four-digit strings in the range between 1100 and 1999.

Number	Reading
2585	Guokteduhát vihttačuođigávccilogivihtta
2.585	"
2 585	"
25700	guoktelogivihttaduhát čiežačuođi
25.700	"
25 700	"
2090350	guokte miljovnna ja ovccilogiduhát golbmačuođivihttalogi
2.090.350	"
2 090 350	"
1000000001	okta miljárda ja okta
3456789012343	golbma njeallje vihtta guhtta čieža gávcci ovcci nolla okta guokte golbma njeallje golbma
34 567 890 123	Golbmaloginjeallje miljárdda ja vihttačuođiguhttalogičieža miljovnna ja gávccičuođiovccilogiduhát čuođiguoktelogigolbma

5.3   Leading zero

Numbers that begin with *0* (zero) followed by a maximum of three digits (the first of which is not *0*) are read *nolla* and then the number as it would be without the *0*. Numbers that begin with *00* (zero zero) followed by a maximum of two digits (the first of which is not *0*) are read *nolla nolla* and then the number as it would be without the *00*. Other digit strings beginning with *0* are read out digit-by-digit.

Number	Reading
0753	nolla čiežačuođivihttalogigolbma
020	nolla guoktelogi
0053	nolla nolla vihttalogigolbma
00753	nolla nolla čieža vihtta golbma
07253	nolla čieža guokte vihtta golbma

5.4   Decimal numbers

Comma is used when writing decimal numbers.

The full number part of the decimal number (the part before comma) is read according to the rules in the section *Full number pronunciation*. If the decimals (the part after comma) are more than three, the decimal part is read as separate digits. Note: A number containing full stop followed by exactly three digits is not read as a decimal number but as a full number, following the rules in the section *Full number pronunciation*.

Number	Reading
16,234	guhttanuppelohkái čuokkis guoktečuođigolbmaloginjeallje
3,1415	golbma čuokkis okta njeallje okta vihtta
1251,04	duhátguoktečuođivihttalogiokta čuokkis nolla njeallje
2,50	guokte čuokkis vihttalogi
2.50	guokte čuokkis vihttalogi
3.141	golbmaduhátčuođinjealljelogiokta

5.5 Currency amounts

The following principles are followed for currency amounts:

- Numbers with zero, one or two decimals preceded or followed by the currency markers *ru, kr, £, $, ¥* or *€* are read as currency amounts.
- Numbers with zero or two decimals followed by the words *kruvnna, kruvnno, kruvnnu, ruvnna, ruvnno, ruvnnu, dollar, yen* or *euro* are read as currency amounts.
- Comma is the only accepted decimal marker.
- The sequence comma followed by hyphen ',-' is read as *ruvdno/ruvnno*
- The decimal part (consisting of two digits) in currency amounts is read as *ja nn evre/evrre*, *ja nn benne*, and *ja nn sente/sentte*.
- If the decimal part is *00* it will not be read.

Example	Reading
kr 20,50	guoktelogi ruvnno ja vihttalogi evrre
kr 20,00	guoktelogi ruvnno
kr 20,-	guoktelogi ruvnno
20,50 kruvnno	guoktelogi ruvnno ja vihttalogi evrre
$15,00	vihttanuppelohkái dollára
15,00£	vihttanuppelohkái punda
€ 200,50	guoktečuođi euro ja vihttalogi sentte
1.000.000 ¥	okta miljovdna yena
$1.314,57	duhát golbmačuođinjealljenuppelohkái dollár ja vihttalogičieža sentte

5.6 Ordinal numbers and date formats

Numbers are read as ordinals only in date formats. The following contexts are valid:

- The number is preceded by a month name/abbreviation. The number is followed by a full stop, space and the abbreviation "b." The number is smaller or equal to 31. The month name/abbreviation may be preceded by a day or an abbreviation for a day.
  Example: ođđajagemánu 3. b.

- The number consists of a day interval preceded by a month name/abbreviation. The interval is followed by the word "beaivvit".
  Example: ođđajagemánu 15.-16. beaivvit

- The number is part of one of the date formats below and occurs in the *dd* part. *yyyy* is a four-digit number number between 1100 and

2099, *yy* is a two-digit number, *mm* is a month number between 1 and 12 and *dd* a day number between 1 and 31.

In type 1 and 2 full stop, and slash may be used as delimiters, in type 3, only hyphen and in type 4 only slash.

In all formats, one or two digits may be used in the *mm* and *dd* part. Zeros may be used in front of numbers below 10.

1. dd/mm/yy
2. dd/mm/yyyy
3. dd-mm yyyy
4. dd/mm yyyy

Valid abbreviations for months: *ođđajagem.*, *ođđajagim.*, *guovvam.*, *njukčam.*, *cuoŋom.*, *miessem.*, *geassem.*, *suoidnem.*, *borgem.*, *čakcam.*, *golggotm.*, *skábmam.* and *juovlam.*

Valid abbreviations for days: *mán, vuos, dis, maŋ, gask, duo, bea, bearj, láv* and *sotn*.

The abbreviations above are only expanded to names of months and days when appearing in correct date contexts.

Expression	Reading	
ođđajagemánu 3. b.	ođđajagemánu goalmmát beaivi	[Not SP]
ođđajagem. 3. b.	ođđajagemánu goalmmát beaivi	[Not SP]
gaskavahkku ođđajagem. 3. b.	gaskavahkku ođđajagimánu goalmmát beaivi	[Not SP]
gaskav. ođđajagem. 3. b.	gaskavahkku ođđajagimánu goalmmát beaivi	[Not SP]
ođđajagemánu 15.-16. beaivvit	ođđajagemánu vihttanuppelogát beaivvis guhttanuppelogát beaivái	[Not SP]
skábmam. 15. b. - njukčam. 15. b.	skábmamánu vihttanuppelogát beaivvis njukčamánu vihttanuppelogát beaivái	
25/12 2004	juovlamánu guoktelogiviđat beaivi guokteduhát njeallje	[Not SP]
25/12/04	"	[Not SP]
25/12/2004	"	[Not SP]
25-12 2004	"	[Not SP]

5.7 Arithmetic operators

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

Expression	Reading	
-12	minus guoktenuppelohkái	
14-2	njealljenuppelohkái sáhcu guokte	[Not SP]
14-2=12	njealljenuppelohkái minus guokte lea guoktenuppelohkái	[Not SP]
+24	plussa guokteloginjeallje	
2+3	guokte plussa golbma	
2+3=5	guokte plussa golbma lea vihtta	
2*3	guokte asterisk golbma	
2*3=6	guokte geardde golbma lea guhtta	
2/3	guokte finjusáhcu golbma	
6/2=3	guhtta juohkit guokte lea golbma	
2:3=6	guokte juohkit golbma lea guhtta	
25%	guoktelogivihtta proseantta	
3,4%	golbma čuokkis njeallje proseantta	

5.8 Mixed digits and letters

If one or more upper-case letters appear within an alphanumeric sequence, the letters are read one by one. The numbers are read according to the examples below.

Expression	Reading
77B184K3	čiežalogičieža B čuođigávcciloginjeallje K golbma
0092B87-B	nolla nolla ovccilogiguokte B gávccilogičieža B
FT2892B87K	F T guoktelogigávcci ovccilogiguokte B gávccilogičieža K
TN12345L5	T N okta guokte golbma njeallje vihtta L vihtta

5.9 Time of day

Numbers denoting time are marked by the abbreviation *dii.* or the word *diibmu* in front of the digits. The Norwegian abbreviations *kl.* and *kl* and

the words *klokken* and *klokka* are also accepted. Either colon or full stop may be used to separate hours, minutes and seconds.

Possible patterns are:

  a. dii. hh.mm or h.mm   kl. hh.mm or h.mm
  b. dii. hh.mm.ss or h.mm.ss   kl. hh.mm.ss or h.mm.ss

Colon may be used instead of full stop and *diibmu* and *klokken/klokka* instead of *dii.* and *kl.* respectively in both patterns. *h* = hour, *m* = minute, *s* = second.

Expression	Reading
dii. 9:00	diibmu ovcci nolla nolla
diibmu 9:30	diibmu ovcci golbmalogi
dii. 13:00	diibmu golbmanuppelohkái nolla nolla
diibmu 12:00	diibmu guokenuppelohkái nolla nolla
dii. 0:00	diibmu nolla nolla nolla
diibmu 20.15.34	diibmu guoktelogi vihttanuppelohkái golbmaloginjeallje
dii. 20.15.34	diibmu guoktelogi vihttanuppelohkái golbmaloginjeallje
kl. 9:00	klokken ovcci nolla nolla
klokken 9:30	klokken ovcci golbmalogi
klokka 13:00	klokka golbmanuppelohkái nolla nolla
20.15.34	guoktelogi vihttanuppelohkái golbmaloginjeallje

5.10 Years

Numbers between 1100 and 1999 are always read as hundreds ("year reading") with the exception of numbers containing decimals. Numbers from 2000 and upwards are read as thousands. Years with case endings :a, :i, :s, :in, :n, :t, :id, :ide, and :iguin are read correctly.

Expression	Reading	
jahki 1988	jahki ovccinuppelotčuođigávccilogigávcci	
jagit 1939-45	jagit ovccinuppelotčuođigolbmalogiovcci gitta njealljelogivihtta	[Not SP]
jagit 1998-2010	jagit ovccinuppelotčuođiovccilogigávcci gitta guokteduhátlogi	[Not SP]
jahki 2004	jahki guokteduhátnjeallje	
1088	duhátgávccilogigávcci	
1900	ovccinuppelotčuođi	[Not SP]

Expression	Reading	
1988	ovccinuppelotčuođigávccilogigávcci	[Not SP]
2000	guokteduhát	
1988,0	duhátovccičuođigávccilogigávcci čuokkis nolla	
1988.32	duhátovccičuođigávccilogigávcci čuokkis golbmalogiguokte	
čakčam. 2004	čakčamánnu guokteduhátnjeallje	[Not SP]

5.11 Phone numbers

In this section the patterns of digits that are recognised as phone numbers are described. In the pronunciation of phone numbers each group of digits is read as a full number (see also *Leading zero* section) with pause between the regional code and the local number, and pauses between groups of numbers.

5.11.1 Ordinary phone numbers

The North Sami system recognizes Norwegian, Swedish and Finnish phone numbers.

Sequences of digits in the following formats are interpreted as phone numbers:

Norwegian:
xx xx xx xx
xxxx xxxx
xxx xx xxx
+xx xx xx xx xx
00 xx xxx xx xx

Examples*:*
22 33 44 55
2233 4455
960 23 123 [Not SP]
+47 22 33 44 55
00 47 22 33 44 55

Swedish:
0x-xx xx xx
0x-xxx xx xx
0x-xxx xxx xx
0xx-xxx xx
0xx-xxx xx

0xx-xx xx xx
0xx-xxx xx xx
0xxx-xxx xx
0xxx-xxx xxx
0xxx-xx xx xx
00xx(0)xx xxx x
00xx-(0)x-xxx xx xx
00xx (0)x-xxx xx xx
+xx-(0)x-xxx xx xx

Examples*:*

08-33 12 31

011-405 31

0121-220 28

+46-(0)8-799 86 19

0049(0)69 116 51

Finnish:
(xx) xxxx
xx xxxx
xxx xxxx
xxxx xxxx
xx xx xxx
xxx xx xxx
xxxx xx xxx
xx xxx xx
xxx xxx xx
xxxx xxx xx
xx xxx xxxx
xxx xxx xxxx
xxx xxx xxxx
xx xxxx xxx
xxx xxxx xxx
xxxx xxxx xxx
xx xxxx xxxx
xxx xxxx xxxx
xxxx xxxx xxxx
xx xx xxxx
xxx xx xxxx
xxxx xx xxxx
xx xxxxx
xxx xxxxx
xxxx xxxxx
xx xxxxxx
xxx xxxxxx
xxxx xxxxxx
xx xxxxxxx
xxx xxxxxxx
xxx xxxxxxx
(xx) xxx xxx
(xxx) xxx xxx
(xxxx) xxx xxx
xx xx xx xx

xx xxx xx xx
xx xxx xx xxx
xx xxx xxx xx
xx xxx xxxxx
xxxx xxx
+x x xx xxx
+xx xx xx xxx
+xxx x xx xxx

Examples*:*

(09) 4568

09 456 8899

0400 4563 889

(016) 123 456

016 45 688

Any of the formats can be preceded with one of the following abbreviations:

tlf.
mob.
mátketlf.
faks
ruktotlf.
bargotlf.
puh
puh.

5.11.2 Special phone numbers

The emergency numbers: 111, 112, 113
Service numbers: 1800 (4-digit numbers)

There are more numbers like this and they are pronounced according to the general rules for pronouncing full numbers.

6 How to change the pronunciation

6.1 User lexicon

Words that are not pronounced correctly by the text-to-speech converter can be entered in the user lexicon (see *User's guide*). When writing transcriptions for entries in the user lexicon to change the way a word is pronounced, one method is to modify the spelling of the word (see section *Spelling incorrectly*) and another is to write a phonetic transcription of the word (see chapter *North Sami phonetic* text). Phonetic transcriptions can also be entered directly in the text, using the *PRN* tag (see *User's guide*).

6.2 Spelling incorrectly

Sometimes the quickest way of changing the pronunciation of the word is to change the spelling of the word directly in the text. Changing a single letter, or adding a hyphen, can often make it sound better.

Example: adoptiona - adopšuvdna

This strategy can also be useful when it comes to foreign words. Try to write the foreign words as they sound in Sami.

Correct spelling	Alternative spelling
light	lajt
photo	foto
date	dejt

7 North Sami phonetic text

The North Sami text-to-speech system uses the North Sami subset of the SAMPA phonetic alphabet (*Speech Assessment Methods Phonetic Alphabet*) with a few exceptions. Many sounds found in Norwegian are also included.

The symbols are written with a space between each phoneme.

Only the symbols listed here may be used in phonetic transcriptions. Symbols not listed here are not valid in phonetic transcriptions and will be ignored if included in the user lexicon or in a *PRN* tag.

7.1 Stress marks

Stress is used to indicate the level of prominence of a syllable in a word (word level stress) or of a word in a sentence (emphasis and reduction). Note that transcriptions may be read with the correct stress even if no stress marks are included, but this happens randomly and is nothing that can be relied on.

In the phonetic transcriptions, primary stress is indicated by the symbol */1/* placed directly after (no space) the accented vowel. Secondary stress is indicated by the symbol */2/*. In compounds and other long words, North Sami has secondary stress on every second syllable after the first secondary stress. Some examples:

bárdni	/ p a:1 r @ ? n i:/
bárdnemánná	/ p aa:1 r n e m a:2 n n a:/
bárdnemánábáideboallu	/ p aa:1 r n e m a:2 n n a: p a:2 j: t e p oa:2 l l u/

7.2 Symbols for the North Sami consonants

The table below lists the phonetic symbols used for the North Sami consonants along with example words and their transcriptions.

Symbol	Word	Phonetic text	Comment
b	doaba	t oa:1 b A	
b:	lábbá	l a:1 b: p a:	
C	čáihni	tS a:1 j C n i:	
C:	duihmmi	t u1 C: m i:	
ds	buozas	p uo1 ds A s	
dS	bulžut	p u1 l: dS u: h t	
ts	cegget	ts e1 g: k e h	

Symbol	Word	Phonetic text	Comment
tS	čakča	tS A1 k: t tS A	
d	dagaldat	t A1 g a l d A h t	
d:	šaddu	S A1 d: t u:	
D	oađđit	oa1 D D i: h	
D:	giđđat	k i1 D: D a: h	
dj	háddjet	h a:1 dj tj e h	
f	fanas	f A1 n a: s	
f:	darffit	t A1 r f: f i: h	
g	logi	l o1 g i	
g:	faggasit	f A1 g: k A s i h	
h	heivet	h e1 j: v e h	
h:	vihtta	v i1 h: t A	
j	jástit	j a:1 s: t i: h	
j:	láiri	l a:1 j: r i:	
k	gelbbolaš	k e1 l p p o l A S	
k:	okto	o1 k: h t o	
l	lihtar	l i1 h: t a: r	
l:	lillut	l i1 l: l u: h	
lj	šiljus	S i1 lj lj u: s	
lj:	Niillas	n i1 lj: lj a: s	
m	máilbmi	m a:1 i2 l @ ? m i	
m:	čaimmas	tS A1 j m: m a: s	
n	noavkut	n oa:1 w: h k u: h	
n:	ivnnit	i1 w n: n i: h	
N	šleŋggodit	S l e1 N k k o D i h	
N:	áŋgir	A:1 N: k i: r	
nj	sutnje	s u1 ? nj e	
nj:	jogŋet	j o1 nj: ? nj e h	
p	bargit	p A1 r @ k i: h	
p:	bippalat	p i1 p: p a: l A h	
r	reivvet	r e1 j v v e h	
r:	šurrat	S u1 r: r a: h	
s	sarrit	s A1 r: r i: h	
s:	lossut	l o1 s: s u: h	
S	šaldi	S A1 l: t i:	
S:	šiššat	S i1 S: S a: h	
t	das	t A1 s	
t:	šattašit	S A1 t: t a: S i h	
T	deatalaš	t ea:1 T A l A S	
tj	bidjat	p i1 tj tj A h	

Symbol	Word	Phonetic text	Comment
tj:	bidjá	p i1 tj: tj a:	
v	vahku	v a1 h: k u:	
v:	čilvvas	tS i1 l v: v a: s	
w	evrret	e1 w r r e h	
w:	rávdat	r a:1 w: t A h	
bb	trøbbel	t r 31 bb e l	
dd	seddel	s e1 dd e l	
gg	ligger	l i1 gg e r	
n=	thorsen	t h u:1 rs n=	syllabic n
rd	verden	v E1 rd n=	retroflex d
rl	farlig	f a:1 rl i	retroflex l
rn	gjerne	j E:1 rn e	retroflex n
rn=	moren	m u:1 rn=	syllabic rn
rs	Lars	l a:1 rs	retroflex s
rt	klart	k l a:1 rt	retroflex t
rt_h	forteller	f o1 rt_h E2 l l e r	aspirated retroflex t

Note that *1* and *2* are stress marks and not part of a consonant or vowel. See section *Stress marks*.

7.2.1 Comments on the phonetic symbols for consonants

7.2.1.1 Retroflexes

The pronunciation of the consonants *t, d, s, l* and *n* is changed when they are orthographically preceded by an *r*. The tip of the tongue is bent backwards against the hard gum and the *r* is not pronounced as a separate sound. These sounds are called supradentals or retroflexes. Retroflexes are not found in North Sami but have been included since they exist in Norwegian and other Scandinavian languages.

7.2.1.2 Aspiration

Aspiration does not occur in North Sami but in Norwegian. The voiceless stops *p, t , k* are typically aspirated in certain positions of the word. That is, they are followed by a "puff of breath" before the vowel. This happens when the stop sound is followed by a stressed vowel and in the beginning of words, including words that are part of compounds. There are no specific allphones for aspirated *p,t,k* in the North Sami system but /p h/, /t h/ and /k h/ before a stressed vowel will have the desired effect. Example: *tø* /t h 3:1/

7.2.1.3 "hll" and "hrr" contexts

Words containing orthographic "hr", "hrr", "hl" and "hll" are transcribed as follows:

h: l	máihlli	m a:1 j h: l i:
w h: l	guvhllit	k u1 w h: l i: h t
h r r	bahrran	p A1 h r r A n
h: r r	skuhrrat	s k u1 h: r r a: h t

7.3 Vowels

The table below lists the phonetic symbols used for the North Sami vowels along with example words and their transcriptions.

Symbol	Word	Phonetic text	Comment
A	agit	A1 g i: h	
a	ahkái	a1 h: k a: j	
a:	bábir	p a:1 b i: r	
aa	áberuvvá	aa1 b e r u2 v: v a:	Used with allegro
e	ebela	e1 b e l A	
e:	ebel	e:1 b e l	
@	olbmot	o1 l @ ? m o h	schwa
i	illu	i1 l: l u	
i:	irgi	i1 r @ k i:	
o	botket	p o1 t: h k e h	
o:	orustit	o:1 r u s t i h	
u	duhát	t u1 h a: t	
u:	oaggut	oa:1 g: k u: h	
ea	ealloravda	ea1 l l o r A2 w: t A	diphthong
ea:	beaivvit	p ea:1 j v v i: h	diphthong
ie	biegga	p ie1 g: k A	diphthong
ie:	diehko	d ie:1 h k o	diphthong
oa	čoamohas	tS oa1 m o h A s	diphthong
oa:	joahcat	j oa:1 h ts A h	diphthong
uo	vuopmi	v uo1 ? m i:	diphthong
uo:	ustitvuođat	u1 s: t i t v uo:2 D A h	diphthong
E	cæsarisma	s E1 s A r i2 s: m a	
E:	lærer	l E:1 r e r	
}	russland	r }1 s l a2 n	
}:	pute	p h }:1 t e	
y	hydrogena	h y1 d r o g e2 n A	

y:	mye	m y:1 e	
3	føderála	f 31 d e r a:2 l A	
3:	lørdag	l 3:1 rd a:2 g	
aU	maurtua	m aU1 r t h }:2 a	diphthong
eI	vei	v eI1	diphthong

7.4 Foreign sounds

Except for the Norwegian sounds in the tables above, there are no other foreign sounds included. However, /T/, /D/ and /w/ are similar to the initial sounds in the English words *think, that* and *what* respectively.

7.5 Phoneme length

Long vowels are marked with colon ':'. Long consonants are marked with colon ':' or by doubling the consonant.

7.6 Glottal stops

A glottal stop, represented by the phonetic symbol */?/*, is a small sound which is often used to separate two words when the second word starts with a stressed vowel. This sound can be inserted in a transcription in order to improve the pronunciation. In Sami, it is also used within words, for example when "tn", "dn", "pm" or "bm" are found in the orthography. The /?/ replaces the /t, d, p, b/ in the transcription.

Examples:

átno   a:1 ? n o
çadni   tS A1 n: ? n i:
albmá   A1 l @ ? m a:
bupme   p u1 ? m e

7.7 Pause

An underscore /_/ in a phonetic transcription generates a small pause.

8   Abbreviations

In the current version of the North Sami text-to-speech system, the abbreviations in the table below are recognised in all contexts. These abbreviations are case-insensitive.

As previously mentioned, there are also abbreviations for the days of the week and the months (see chapter *Ordinal numbers and date formats*).

Abbreviation	Reading
čilg.	čilgehus
fuom.	fuomáš
ng.	nu gohččoduvvon
tel.	telefonnummár
tlf.	telefonnummár
rievd.	rievdadus
oá.	ovdal áigelogu
osd.	ossodat
ee.	earret eará
ea.ea.	earret eará
čuj.	čujuhus
omd.	ovdamearkka dihte
vrd.	veardit
gč.	geahča
dii.	diibmu
dhj.	dahje
jna.	ja nu ain
jnv.	ja nu viidásut
dearv.	dearvvuođat
dr.	doavttir

Some abbreviations representing units of measurement and measures of capacity are only expanded after digits. Abbreviations connected to telephony are only expanded in front of digits.

Abbreviation	Reading
20 ml	guoktelogi millilihtara
25 cl	guoktelogivihtta centilihtara
30 dl	golbmalogi decilihtara
40 mm	njealljelogi millimehter
50 cm	vihttalogi centimehter
60 dm	guhttalogi decimehter
70 kg	čiežalogi kilo

Abbreviation	Reading
tel 32 12 34 56	telefonnummár golbmalogiguokte (pause) guoktenuppelohkái (pause) golbmaloginjeallje (pause) vihttalogiguhtta
mob 245 35 043	mobiila guoktečuođinjeallelogivihtta (pause) golbmalogivihtta (pause) nolla njealljelogigolbma

9 Web-addresses and email

Web-addresses and email-addresses are read as follows:

- *www* is read as three normal *v*'s spelled letter by letter.
- Full stops '.' are read as *čuokkis*, hyphens '-' as *sáhcu*, underscores '_' as *deattuhusmearka*, and slashes '/' as *finjusáhcu*.
- The @ is read *smarfe-a*.
- Words/strings (including *org*, *com* and *edu*) are pronounced according to the normal rules of pronunciation in the system and in accordance with the lexicon.

Expression	*Reading*
www.acapela-group.com	v v v čuokkis acapela sáhcu group čuokkis com
support@acapela-group.com	support smarfe-a acapela sáhcu group čuokkis com
http://www.acapela-group.com	h t t p duppalčuokkis finjusáhcu finjusáhcu v v v čuokkis acapela sáhcu group com
www.samediggi.no	v v v čuokkis samediggi čuokkis no
john_smith@yahoo.com	john deattuhusmearka smith smarfe-a yahoo čuokkis com