



Language Manual

# Belgian Dutch

Sofie and Els

Language Manual  
Belgian Dutch  
Sofie and Else  
1 January 2005, modified 11 april 2007

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](http://www.acapela-group.com)

© Copyright Acapela Group 2005. All rights reserved.

## List of contents

1	General.....	4
2	Letters in orthographic text.....	5
3	Punctuation characters .....	6
3.1	Comma, colon and semicolon.....	6
3.2	Quotation marks.....	6
3.3	Full stop .....	6
3.4	Question mark.....	6
3.5	Exclamation mark .....	6
3.6	Parentheses, brackets and braces.....	6
4	Other non-alphanumeric characters .....	7
4.1	Non-punctuation characters .....	7
4.2	The <sup>2</sup> and <sup>3</sup> signs.....	7
4.3	Symbols whose pronunciation varies depending on the context.....	8
4.3.1	Hyphen.....	8
4.3.2	Asterisk.....	8
5	Number processing .....	9
5.1	Full number pronunciation.....	9
5.2	Leading zero .....	10
5.3	Decimal numbers .....	10
5.4	Currency amounts .....	10
5.5	Ordinal numbers.....	11
5.6	Arithmetic operators .....	11
5.7	Mixed digits and letters.....	11
5.8	Time of day .....	11
5.9	Year.....	12
5.10	Dates .....	12
5.11	Phone numbers.....	13
5.11.1	Ordinary phone numbers.....	13
5.11.2	International phone numbers.....	13
6	How to change pronunciation errors .....	14
7	Belgian Dutch Phonetic Text .....	15
7.1	Consonants.....	15
7.1.1	Symbols for the Belgian Dutch consonants .....	15
7.2	Vowels .....	16
7.2.1	Symbols for the Belgian Dutch vowels.....	16
7.3	Lexical stress.....	16
7.4	Glottal stops .....	16
7.5	Pause .....	16
8	Abbreviations.....	17
9	Web-addresses and email.....	18

# 1 General

This document discusses certain aspects of text-to-speech processing for the Belgian Dutch 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 voice Sofie and the High Density voice Els.

## 2 Letters in orthographic text

Characters from A-Z and a-z may constitute a word. Certain other characters are also considered as letters, notably those used as letters in other European languages, i.e. “ñ, ò, à, ç, é”. These letters are not pronounced as in their native languages though, they are pronounced as regular “n, o, a, c, e” when occurring in a word.

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 5) and in abbreviations (see chapter 8).

### *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 < ! > behaves 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.

Symbol	Reading
/	Slash
+	plus
\$	Dollar
£	Pond
€	Euro
¥	Yen
<	kleiner dan
>	groter dan
%	Procent
^	accent circonflexe
	rechte streep
~	Tilde
@	Apestaart
=	Gelijkheidsteken
<sup>2</sup>	see below
<sup>3</sup>	see below
-	see below
*	see below

**Table 1** Non-punctuation characters

### 4.2 The <sup>2</sup> and <sup>3</sup> signs

The reading of expressions with <sup>2</sup> and <sup>3</sup> is:

Expression	Reading
mm <sup>2</sup>	vierkante millimeter
cm <sup>2</sup>	vierkante centimeter
m <sup>2</sup>	vierkante meter
km <sup>2</sup>	vierkante kilometer
mm <sup>3</sup>	kubieke millimeter
cm <sup>3</sup>	kubieke centimeter
m <sup>3</sup>	kubieke meter
km <sup>3</sup>	kubieke kilometer

### 4.3 Symbols whose pronunciation varies depending on the context

#### 4.3.1 Hyphen

A hyphen < - > is pronounced "minus" if followed by a digit. In certain date formats, in between days or years, the hyphen is pronounced "to". In other cases the hyphen is never pronounced.

Expression	Reading
44-3	vierenveertig min drie
15-20 Oktober	vijftien tot twintig oktober
6-10 Nov	zes tot tien november
1998-2004	negentienhonderd achtennegentig tot tweeduizendenvier
02-02-2002	twee februari tweeduizend en twee
zuidoost-azië	zuidoost azië

#### 4.3.2 Asterisk

Asterisk < \* > is pronounced "times" if enclosed by digits. In other cases it is pronounced "asterisk".

Expression	Reading
2*3	twee maal drie
*bc	asterisk be se

## 5 Number processing

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:

Full number pronunciation **Erreur ! Source du renvoi introuvable.**

Leading zero

Decimal numbers

Currency amounts

Ordinal numbers

Arithmetic operators

Mixed digits and letters

Time of day

Year

Dates

Phone numbers

### 5.1 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 point (not full stop). 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 9999999999 (eleven digits). Numbers higher than this are read as separate digits.

Number	Reading
2580	twee duizend vijfhonderd tachtig
2 580	twee duizend vijfhonderd tachtig
2.580	twee duizend vijfhonderd tachtig
25800	vijfentwintig duizend achthonderd
25 800	vijfentwintig duizend achthonderd
25.800	vijfentwintig duizend achthonderd
2580350	twee miljoen vijfhonderd tachtig duizend driehonderd vijftig
2 580 350	twee miljoen vijfhonderd tachtig duizend driehonderd vijftig
2.580.350	twee miljoen vijfhonderd tachtig duizend driehonderd vijftig
1000000000	één miljard
123456789012	één twee drie vier vijf zes zeven acht negen nul één twee

## 5.2 Leading zero

Numbers that begin with 0 (zero) are read digit by digit.

Number	Reading
09253	nul negen duizend tweehonderd driënvijftig
020	nul twintig

## 5.3 Decimal numbers

Comma or full stop may be used when writing decimal numbers.

The full number part of the decimal number (the part before comma or full stop) is read according to the rules in 5.1. The decimals (the part after comma or full stop) are read as separate digits if there are more than 3 digits after the comma. Note: A number containing a point followed by exactly three digits is not read as a decimal number but as a full number, following the rules in 5.1.

Number	Reading
16,234	zestien komma tweehonderd vierendertig
3,1415	drie komma één vier één vijf
1251,04	duizend tweehonderd éénenvijftig komma nul vier
1.251,04	duizend tweehonderd éénenvijftig komma nul vier
2.50	twee punt vijftig
2,50	twee komma vijftig
3,141	drie komma honderd eenenveertig

## 5.4 Currency amounts

The following principles are followed for currency amounts:

- Numbers with zero or two decimal places preceded or followed by the currency markers £, \$, ¥ or € are read as monetary amounts.
- Numbers with zero or two decimal places followed by the words "pounds", "dollars", "yen" or "euros" (singular or plural) are read as monetary amounts.
- Accepted decimal markers are comma and full stop.
- No spaces are allowed in the number.
- The decimal part (consisting of two digits) in monetary amounts is read as "and nn pence" and "and nn cents".
- If the decimal part is "00" it will not be read.

Example	Reading
\$15,00.	vijftien dollar
15,00£.	vijftien pond
15,00 euros.	Vijftien euro
€ 200,50	tweehonderd euro en vijftig cent
1.000.000 ¥	één miljoen yen

There is also the possibility of writing large amounts as follows:

\$ 1 miljoen	één miljoen dollar
--------------	--------------------

## 5.5 Ordinal numbers

Numbers are read as ordinals in the following case:

- The number is followed by "de, ste". Examples: 1ste, 2de .

**Valid abbreviations for months:** Jan, Feb, Mar, Apr, Jun, Jul, Aug, Sep, Okt, Nov and Dec.

**Valid abbreviations for days:** Maa, Din, Woe, Don, Vrij, Zat and Zon.

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

## 5.6 Arithmetic operators

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

Expression	Reading
-12	min twaalf
+24	plus vierentwintig
2*3	twee maal drie
2/3	twee gedeeld door drie
25%	vijfentwintig procent
3.4%	drie punt vier procent

## 5.7 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.

Examples:

Expression	Reading
77B84Z3	zevenenzeventig be vierentachtig zed drie
0092B87-B	nul nul tweeënnegentig be zevenentachtig be

## 5.8 Time of day

The colon is used to separate hours, minutes and seconds.

Possible patterns are:

- a) hh:mm (or h:mm)
- b) hh:mm:ss (or h:mm:ss)
- c) hh:mm'ss" (or h:mm'ss") ex 12:30'45"
- d) hh:mm u ex 15:25 u
- e) hh.mm u ex 15.25 u

h = hour, m = minute, s = second.

In pattern a): If the "mm"-part is equal to "00", this part will not be read.

Example: 9:00  
13:00

In pattern b): An "en" will be inserted before the "ss"-part, and "seconds" will be added after it. If the "ss"-part is equal to "00", this part will not be read.

Pattern (c) follows the rules for pattern (b).

## 5.9 Year

Numbers between 1100 and 2000 are always read as hundreds (year reading) with the exception of numbers containing decimals.

Expression	Reading
1988	negentienhonderd achtentachtig
1939-45	negentienhonderd negenendertig tot vijfenveertig
1088	duizend achtentachtig
1988,0	duizend negenhonderd achtentachtig komma nul
1988,32	duizend negenhonderd achtentachtig komma tweeëndertig
September 1999	september negentienhonderd negenennegentig

## 5.10 Dates

The valid formats for dates are:

- 1.dd-mm-yyyy, dd.mm.yyyy, and dd/mm/yyyy
- 2.dd-mm-yy, dd.mm.yy, and dd/mm/yy

“yyyy” is a four-digit number, “yy” is a two-digit number, “mm” is a month number between 1 and 12 and “dd” a day number between 1 and 31.

Hyphen, full stop and slash may be used as delimiters.

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.

Examples of valid formats and their readings:

Type 1: dd-mm-yyyy, dd.mm.yyyy, and dd/mm/yyyy

10-02-2003	or	10-2-2003	tien februari tweeduizend en drie
10.02.2003	or	10.2.2003	tien februari tweeduizend en drie
10/02/2003	or	10/2/2003	tien februari tweeduizend en drie

Type 2: dd-mm-yy, dd.mm.yy, and dd/mm/yy

10-02-03	or	10-2-03	tien februari tweeduizend en drie
10.02.03	or	10.2.03	tien februari tweeduizend en drie
10/02/03	or	10/2/03	tien februari tweeduizend en drie

Ranges of days and years are also supported.

Examples:

1998-1999	negentienhonderd achtennegentig negentienhonderd negenennegentig
1939-45	negentienhonderd negenendertig tot vijfenveertig
2002/3	tweeduizend en twee tot drie
14-15 Januari	veertien tot vijftien januari
Oktober 19-20	oktober negentien tot twintig

Other possible formats include :

- Maandag, 15 Januari (with or without the comma) Maandag vijftien januari
- 30 April 1999 dertig april negentienhonderd negenennegentig
- April 30 1999 april dertig negentienhonderd negenennegentig



## 6 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). In this lexicon, the user enters a phonetic transcription of the word (see chapter 7). Phonetic translations can also be entered directly in the text, using a PRN-tag (see User's guide).

## 7 Belgian Dutch Phonetic Text

The Belgian Dutch text-to-speech system uses the Dutch subset of the SAMPA phonetic alphabet (Speech Assessment Methods Phonetic Alphabet). The symbols are written with a space between each phoneme.

Only SAMPA 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 Consonants

#### 7.1.1 Symbols for the Belgian Dutch consonants

Symbol	Word	Phonetic text	Comment
p	Pad	\p A t\	
t	Tak	\t A k\	
tj	potje	\p O tj @\	
k	Kat	\k A t\	
b	Bad	\b A t\	
d	Dak	\d A k\	
dj	Djinn	\dj I n\	
g	zakdoek	\z A g d u k\	
f	fiets	\f i t s\	
s	sap	\s A p\	
S	Sjall	\S a l\	
x	lach	\l A x\	
v	Vat	\v A t\	
z	zat	\z A t\	
Z	plantage	\p l A t a Z @\	
G	regen	\r e G @ n\	
h	Huis	\h 9y s\	
w	sneeuwen	\s n e w @\	
j	aaien	\a j @\	
l	alle	\A l @\	
r	haar	\h a R\	
m	mat	\m A t\	
n	Nat	\n A t\	
N	lang	\l A N\	
nj	anjer	\A nj @ r\	

Table 2 Belgian Dutch consonants

## 7.2 Vowels

### 7.2.1 Symbols for the Belgian Dutch vowels

Symbol	Word	Phonetic text	Comment
ɪ	bid	\b ɪ t\	
E	bed	\b E t\	
Y	Buts	\b Y t s\	
O	Bos	\b O s\	
@	Rede	\r e d @\	
i	bied	\b i t\	
e	beet	\b e t\	
y	buut	\b y t\	
2	Beuk	\b 2 k\	
a	Baat	\b a t\	
u	boek	\b u k\	
o	Boot	\b o t\	
Ei	bijd	\b Ei t\	
9y	Buit	\b 9y t\	
Au	bout	\b Au t\	
E~	Timbre	\t E~ b R @\	
A~	Chanson	\S A~ s O~\	
O~	Bonbon	\b O~ b O~\	
A	Bak	\b A k\	
Oe	Ferule	\f R Oe l @\	

Table 3 Belgian Dutch vowels

### 7.3 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. For the correct pronunciation of a word, it is important to include the symbol marking the word stress.

In the phonetic transcriptions the word stress is indicated by the symbol "1" placed directly after the stressed vowel (with no space between the vowel symbol and the stress symbol).

### 7.4 Glottal stops

A glottal stop, represented by the phonetic symbol /ʔ/, 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.

### 7.5 Pause

An underscore < \_ > in a phonetic transcription generates a small pause.

## 8 Abbreviations

In the current version of the Belgian Dutch text-to-speech system, the abbreviations in table 4 below are recognised in all contexts. These abbreviations are mostly case-insensitive and require no full stop in order to be recognised as an abbreviation.

As previously mentioned, there are also abbreviations for the days of the week and the months.

Abbreviation	Reading
DM	dOjtSmArk
Hz	hErts
Kcal	kilokAlori
MHz	meGahErts
MWh	megawAtp@ryr
Mme	mAdAm@
tel	telefon
µV	mikrovOlt
µg	mikroGrAm
µm	mikromet@r
µs	mikros@kOnd@
°C	GratsElsjYs
°F	Gratfar@nhEit
°K	GratkElvln
Sr	senjOr
Jr	jYnjOr
km/h	kilomet@rp@ryr
mg	miliGrAm
kbit/s	kilobltp@rs@kOnd@
kbit	kilobltp
kVA	kilovOltAmper@
kbyte	kilobAjt
ml	mililit@r
lm	lymEn
mA	miliAmper
mV	milivOlt
mW	miliwAt
mbar	milibar
tem	tOt@nmEt
Mt	megatOn
Wh	wAtp@ryr
tav	t@rAtEntsivAn
mbt	mEtB@trEkINtOt
dwz	dAtwllzEG@n
tem	tOt@nmEt

**Table 4** Abbreviations

## 9 Web-addresses and email

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

- "www" is read as three w's spelled letter by letter.
- Full stops are read as "punt", hyphens as "min", underscore ("\_") as "onderstreep", slash ("/") as "slash".
- "us, uk, fr" and all the other abbreviations for countries are spelled out letter by letter.
- The "@" is read "at" .
- 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.

String

Reading

[www.babeltech.com](http://www.babeltech.com)

<http://www.babeltech.com>

[smith@yahoo.us](mailto:smith@yahoo.us)

[jane\\_smith@yahoo.us](mailto:jane_smith@yahoo.us)