



Technical Report

Subject: Transcription conventions for Norwegian

Version: 1.5

Date: 15.08.00	Author: Helga Nygaard Øverland
Project: Ldb	Confidentiality: NST/VILT

TRANSCRIPTION CONVENTIONS FOR LEXICAL DATABASES

Language: Norwegian
Revision: 1.05
Author: Helga Nygaard Øverland
(based on Bente Henrikka Moxness' v.1.1 – 1.4)
Date: 2000.08.15

Table of Contents	2
0 INTRODUCTION	4
0.1 Intended use	4
0.2 The terminology	4
0.3 General rules	5
1 SYLLABLE STRUCTURE	5
1.1 Syllable boundaries and MOP	5
1.1.1 MOP in decompounded words	5
1.1.2 MOP and acronyms	7
2 CONSONANTS	7
2.1 The velar nasal	7
2.2 The retroflexes	8
2.3 Omission of /d/, /t/ and /g/	9
2.4 Syllabification	10
2.5 The postalveolar fricative (/S/)	12
2.5.1 A phoneme of its own	12
2.5.2 <rs> → /S/	12
2.5.3 <sl> → /Sl/	12
2.5.4 In separate morphemes and in compounds	13
2.6 The realization of word-initial <tj>, <kj>, <ki>	13
2.6.1 The affricate /t&S/	14
2.7 The realization of the word-initial <gi>, <gy>, <gei> and <gøy>	14
2.8 The realization of the word-intitial <sk>	14
3 VOWELS	15
3.1 /U/ or /O/?	15
3.2 Vowels in unstressed position	15
3.3 The semi-vowel /j/	16

4	PROSODY	16
4.1	Stress	16
4.1.1	Primary stress	16
4.1.2	Secondary stress	17
4.1.3	Stress and tonemicity in foreign words	18
4.2	Length	19
4.2.1	Length in primary stressed syllables	19
4.2.2	Length in secondary stressed syllables	19
4.2.3	Length in unstressed syllables	19
4.3	Tonemicity	20
5	OTHER SPECIFIC TRANSCR. CONVENTIONS	20
5.1	<veg> and <høg>	20
5.2	Foreign words ending on <-or>	20
5.3	Words in “nynorsk”	21
5.4	Words in “riksmål”	21
5.5	Multiple word utterances (phrases)	21
5.6	<y> in unstressed position in names	21
5.7	Assimilation vs. phonemic transcription	21
5.7.1	Devoicing	21
5.7.2	Transitional changes due to compounding	22
5.8	Phonological variants	22
5.9	Acronyms and abbreviations	23
6	Diacritics and suprasegmentals	23

Appendices

Appendix 1	Possible onsets in Norwegian	23
Appendix 2	Decompounding	25
Appendix 3	Transcription of foreign words	31
Appendix 4	Transcription of names	32
Appendix 5	Coding of transcriptions	37
Appendix 6	Transcription of multiple word phrases	40
Appendix 7	The fields of the transcription lexicon	42
Appendix 8	Possible codas in Norwegian	43

0. INTRODUCTION

01. Intended use

This document is outlining some transcription conventions for the use of L&H+ applied to the chosen spoken language variant of Norwegian, Ålesund bymål. The document is intended for instructional use for those involved in the transcription of Norwegian at Nordisk Språkteknologi AS, and it is based loosely on the phonetic/phonological principles in Norwegian transcription traditions. However, the main foundation for the conventions, is the experience made during the transcription of the initial pronunciation lexicon for Norwegian done in PhonEd by a group of speech technicians at NST. These experiences also include a tape recording of a 24-year old female native speaker of Ålesund bymål. We also bear in mind that the pronunciation lexicon we are creating will sound like a person from Ålesund reading the written Norwegian language bokmål.

As these transcriptions constitute the backbone for all applications, they should be as general as possible. They should have a high degree of distinctness.

Some conventions are determined in the PhonTable v1.08 (the subset of L&H+ symbols for Norwegian, which are considered phonemes, which are allophones, etc.) Other conventions have been established as we went along. The recording of the native speaker of the chosen language variant clarified some issues. However, since speakers are seldom consistent and since we have to establish a norm, the pronunciations by the native speaker are not considered complete defaults, and we sometimes decide to transcribe more in accordance with general standardized bokmål pronunciation.

Sections 1 and 2 describe consonants and vowels, while section 3 deals with prosody. Section 4 takes care of the syllable structure, while section 5 describes some special conventions. Finally, applied diacritics and suprasegmentals are listed in section 6.

For reference we use Arne Vanvik: A Norwegian Pronunciation Dictionary (Norsk uttaleordbok). Reference is applied for the sake of consistency in cases where neither conventions nor the lexicographers' judgement as native speakers is adequate. We choose the first example and avoid the alternatives in brackets.

0.2 The terminology

A list of short definitions of some of the most important terms regarding transcription and syllables used in this document follows below. Some might already be acquainted with the terms, while others with less phonetic and phonological training might find them useful.

Transcription: The method of writing down speech sounds in a systematic and consistent way.

Phonemic transcription: The units to be transcribed are those which have a linguistic function, i.e. the phonemes (as opposed to a phonetic transcription, which transcribes units according to their articulatory/auditory identity, regardless of their function.).

E.g.: phonemic transcription /pe|n/, phonetic transcription [pH][[ε]]v]

Allophonic transcription: Adds functional phonetic details. E.g.: syllabic consonants (see definition below).

Syllable: A unit of pronunciation typically larger than a single sound and smaller than a word. The notion of syllable is very real to native speakers. A precise definition of the syllable does not exist. There are several theories in both phonetics and phonology. A phonological approach is taken in this document, since we focus on the way sounds combine in an individual language to produce typical syllables. Two classes of sounds emerge: those sounds which can occur on their own, or at the center of a syllable, and those sounds which cannot occur on their own, but rather at the edges of a syllable. The former are referred to as vowels or centers and the latter as consonants or edges. All syllables must have a vowel/center, while consonants/edges need not always be present.

Onset: The opening segment of a syllable, a consonant/edge.

Nucleus: The obligatory center of a syllable, a vowel/center.

Coda: The closing segment of a syllable, a consonant/edge.

Rhyme: The nucleus and coda together make up the rhyme of the syllable.

E.g.: in the monosyllabic word *dog*, phonemically: /dog/, the onset is /d/ whereas the /og/ makes up the rhyme (/o/) acts as nucleus and /g/ as coda).

Syllabic consonant: In Norwegian, a nasal or lateral consonant which in exceptional cases function as the vowel/center of a syllable. In Norwegian this occurs as a result of a phonological process of surface reduction/elimination of an underlying vowel.

0.3 General rules

The transcription should be as phonemic as possible and in accordance with the Norwegian subset of L&H+ (current: PhonTable v.1.09). Therefore, <tenkt> is transcribed /tEnKkt/ (Cf. 1.2 below) with a velar plosive following the velar nasal, even though the velar plosive is hardly audible. In some cases we make use of allophonic transcriptions.

Some special adaptations have to be considered and explained. The explanations are outlined below.

1. SYLLABLE STRUCTURE

1.1 Syllable boundaries and MOP

Syllabic boundaries follow a moderate Maximal Onset Principle (MOP), which means that Norwegian phonotax is largely followed in order to avoid impossible consonant clusters in the onsets. This means that the MOP we apply in our transcriptions allows all onsets that can be found word-initially (i.e. at the beginning of a word) in Norwegian. (See Appendix 1 for a list of possible Norwegian onsets).

1.1.1 MOP in decompounded words

Morphological boundaries preside over MOP in compounds. (See Appendix 2: “Decompounding (NO)” v.3.05, Aug. 2000). The linking morpheme <s> is always added to the syllable preceding the compound boundary.

Example:

lærerutdanning (E: teacher training)	lærer + utdanning	2'1@:.r6Er6.u0:t.'2dA.nInK
dagsverk (Eng: day's work)	dag+s+verk	'dA:gs.'2VEr6k

In decompounded words each constituent is transcribed individually. Also, each part keeps the vowel quality of the individual word even though there may be a reduction of the vowel in the first constituent, cf. transcription above. For reduction of vowels, cf. 5.7).

In words where the last consonant in the first component is the same as the first consonant in the following component, this consonant is transcribed in both components. This also applies to compounds with a linking <s> when the second component starts with an /s/ or an /S/:

Examples:

<i>Orthography</i>	<i>Decompounding</i>	<i>Transcription</i>
lovverket (Eng: laws)	lov + verket	2'lo:V.'2VEr6.kE
velferdssamfunnet (Eng: welfare society)	velferd + s + samfunnet	2'VEl.fEd+S.sAm.'2fU0.nE
statssekretær (E: secretary of state)	stat + s + sekretær	'stA:ts.sE.kr6E.'2t@:r6
lyssky (Eng: fishy)	lys + sky	2'ly:s.'2Sy:

In words where the last consonant in the first component and the first consonant in the next component constitute a retroflex, the word is transcribed without a retroflex.

Examples:

amatørteater (E: amateur theatre): amatør + teater (rt→t+)	A.mA.'te+:r6.tE.'2A:.tEr6
bardisk (E: bar) bar + disk (rd→d+)	'bA:r6.'2dIsk
teaterleder (E: theatre manager) teater + leder (rl→l+)	tE.'A:.tEr6.'2le:.dEr6
gårdnummer (E: block number) gård + nummer (rn→n+)	2'go:r6.'2nU.mEr6

When a retroflex is followed by t, d, l, s or n, these will usually also become retroflexed. This is not transcribed in the case of retroflexation across compound boundaries:

Examples:

jernteppe (Eng: iron curtain)	jern + teppe	2'j@:n+.'2tE.pE
jerntrapp (Eng: iron ladder)	jern + trapp	2'j@:n+.'2tr6Ap

The linking morpheme <s> is added to the first component (cf. above):

Examples:

ventilasjonsanlegget (E: ventilation system)	ventilasjon+s+anlegget	VEn.tI.lA.'Su:ns.An.'2IE.gE
verdenskart (Eng: world map)	verden + s + kart	'VE.d+n+%s.'2kAt+

A linking <s> preceded by a retroflex becomes the postalveolar fricative /S/:

Examples:

velferdspermisjon (E: welfare leave)	velferd + s + permisjon	2'VEl.fEd+S.pEr6.ml.'2Su:n
--------------------------------------	-------------------------	----------------------------

opprørsbevegelsen (E: revolt movement)	opprør + s + bevegelsen	2'O.pr6e+:S.bE.'2Ve:.gl%).sn%)
--	-------------------------	--------------------------------

But when there is a syllabification between the retroflex and the <s>, the <s> will not become postalveolar:

ordensmenneske (Eng: methodical person)	'O.d+n+%)s.'2mE.nE.skE (/s/ is not influenced)
---	--

1.1.2 *MOP in acronyms*

When acronyms are pronounced letter by letter, they are also transcribed letter by letter, i.e. without the MOP. (For secondary stress before the main stress – cf. 2.2.2. and 2.3.)

Example:

NRK-program	NRK-+-program	'2En.Er6.'ko:.pr6U.'2gr6Am
-------------	---------------	----------------------------

Acronyms pronounced as a word (as in NATO), follow the general rules for MOP:

Example:

NATO	'nA:.tU
------	---------

In compounds containing an acronym the acronym is treated as a regular word and follows the general rules:

Example:

OPEC-landene:	OPEC +++ landene	'u:.pEk.'2lA.nn%).E
---------------	------------------	---------------------

2. CONSONANTS

2.1 The velar nasal

The velar nasal occurs only in coda position and is a result of the cluster <ng>. Norwegian phonotax disallows this sound from occurring initially in a syllable and morpheme. The reason why it is mentioned here as an adaptation, is that in German, a language in which the velar nasal follows the same phonotactical rules as Norwegian, the sound has been transcribed in the onset.

Example:

lange (Eng: long (plural))	2'lAnK.E
----------------------------	----------

A nasal in final position of a lexeme or a morpheme followed by the plosive <k> will make up a velar nasal followed by /k/. The same process occurs to a <g> when it is not part of the same morpheme and is pronounced /g/:

Example:

angå (Eng: concern)	'AnK.'2go:
---------------------	------------

<i>but</i> : angikk (Eng: concerned)	'An.'2jlk (<g> in front of <i> and <y> makes /j/)
anke (Eng: appeal)	2'AnK.kE (morpheme internal)

Also <nd> followed by a <k> is pronounced nKk:

håndkle (Eng: towel)	2'hOnK.kIE
----------------------	------------

In decomposed words the lemmas are transcribed separately and the velar nasal is not reflected in the transcription (cf. 5.7):

vannglass (E: drinking glass)	vann + glass	2'VAn.'2glAs
båndkassett (E: tape cassette)	band + kassett	2'bOn.kA.'2sEt

A nasal in front of <s> in French loan words may be transcribed as both a velar nasal and a plain nasal. Our first choice is here the velar nasal (cf. Vanvik).

finanser (Eng: finances)	fl.'nAnK.sEr6 (first variant)
	fl.'nAn.sEr6 (second variant)
balanse (Eng: balance)	bA.'lAnK.sE (first variant)
	bA.'lAn.sE (second variant)
pensjonist (Eng: retired person)	pAnK.SU.'nIst (first variant)
	pEn.SU.'nIst (second variant)

Words containing <gn> in coda position are in many instances pronounced with a velar nasal plus <n>. For reference, see Vanvik.

vogn (Eng: wagon)	'VOnKn
ligning (Eng: equation)	2'lInK.nInK
<i>but</i> :	
bygning (Eng: building)	2'bY.gnInK
bevilgning (Eng: grant)	bE.'VII.gnInK

2.2 The retroflexes

The retroflex plosives, the retroflex nasal and the retroflex lateral occur in onset position in the transcriptions although this clashes with Norwegian phonotax. These sounds always occur in word-internal or word-final positions and are underlyingly represented by combinations of alveolars with /r/:

/rt/ → /t+/
 /rd/ → /d+/
 /rn/ → /n+/
 /rl/ → /l+/

However, with respect to VOT (voice onset time) and duration the retroflexes are very similar to alveolar plosives, nasals and laterals in onset positions, and are thus treated along the same lines as the alveolar sounds for our purposes which means that (t+), (d+), (n+) and (l+) can also serve as onsets, in front of syllable nuclei (vowels and syllabic consonants).

Examples:

burde (Eng: ought)	2'bU0.d+E
fortelle (Eng: tell)	fO.'t+E.IE
hjerne (Eng: brain)	2'j@:.n+E
forlagt (Eng: misplaced)	fO.'l+Agt

In compounds the coda in the first constituent and the onset in the next are transcribed as if they were separate words:

Examples:

vinternatt (Eng: winter night)	(vinter+natt)	2'VIn.tEr6.'2nAt
bardisk (Eng: bar)	(bar+disk)	'bA:r6.'2dIsk

When a retroflex is followed by a t, d, l, s or n also these become retroflex:

Example:

verden (Eng: the world) (syllabeled <n>)	'VE.d+n+%)
barnlig (Eng: like a child)	2'bA:n+.l+I
ordentlig (Eng: proper)	O.d+n+%)t+.l+I

But when there is a syllabification between the retroflex and the <s>, the <s> will not become postalveolar:

ordensmenneske (Eng: methodical person)	'O.d+n+%)s.'2mE.nE.skE (/s/ is not influenced)
---	---

Influenced by a preceding retroflex, an <s> will become postalveolar /S/. This also implies genitive-s. But a following consonant will remain uninfluenced.

Examples:

verdslig (Eng: secular)		2'VEd+.SII
fartssynder (Eng: speeder)	fart+s+synder	'fAt+S.'2sYn.dEr6

2.3 Omission of /d/, /t/ and /g/

The plosive /d/ in morpheme-final position is in some cases not pronounced. However, if the underlying /d/ is likely to be pronounced in derivations or inflexions of the morpheme, the /d/ is to be transcribed.

An exception is the adjective <god> (Eng: good) which will always be pronounced /gu:/ (without the final /d/).

Examples:

tid: (Eng: time)	'ti:d	The most frequent pronunciation is: /'ti:/ but when inflecting the word, the /d/ reappears: /'ti:.dn%) (Eng: the time), <i>but</i> /'ti:.A/
arbeid (Eng: work)	2'Ar6.bE&Id	The most frequent pronunciation is: /2'Ar6.bE&I/, but in derivations of the word, the /d/ reappears: /2'Ar6.bE&Ids.mU.'2r6A:l (Eng: work ethics/morale)
god (Eng: good)	'gu:	

The plosive /t/ is never pronounced in singular definite forms of neuter nouns, which have the ending <et>. In singular definite genitive form it is pronounced, however.

Some of these nouns may be homographs to past tense of weak verbs, which also have the ending <et>, and the /t/ is then pronounced in these cases. These verbs have toneme 2, while the corresponding noun has toneme 1. (For toneme, see 2.1):

Examples:

landet (Eng: the country)	'lA.nE	noun, sg. def. toneme 1 (is homograph to a verb)
landet (Eng: landed)	2'lA.nEt	verb, past tense, toneme 2 (is homograph to a noun)
laget (Eng: the team)	'lA:.gE	noun, sg. def. toneme 1 (is homograph to a verb)
laget (Eng: made)	2'lA:.gEt	verb, past tense, toneme 2 (is homograph to a noun)
barnet (Eng: the child)	'bA:.n+E	noun, sg. def. toneme 1 (is not homograph to a verb)
barnets (Eng: the child's)	'bA:.n+Ets	noun, sg. def. genitive
lusket (E: sneaked)	2'lU0.skEt	verb, past tense, toneme 2 (not homograph to a noun)

The plosive /g/ distinguishes the homographs *følge* (verb, Eng: follow) and *følge* (noun, Eng: result). The plosive /d/ distinguishes the homographs *gode* (adjective, Eng: good, pl.) and *gode* (noun, Eng: benefit).

Examples:

følge (Eng: follow) – verb	2'fE+.lE (withoug /g/)
følge (Eng: result) – noun	2'fE+l.gE (with /g/)
gode (Eng: good) – adjective	2'gu:.E (without /d/)
gode (Eng: benefit) – noun	2'gu:.dE (with /d/)

Monosyllabic words ending in <rd> originating from old Norse <ρΔ> do not get retroflexes. The final <d> becomes mute, as in *nord*, *fjord*, *jord*, *bord*, *ord*, *hard*, *gård* - also in derivations and compounds. No statistics is available to help us estimate the proportion of these words in a text corpus.

Examples:

hard (Eng: hard)	'hA:r6
hardfør (tough)	2'hA:r6.'2fe+:r6

2.4 Syllabification

In Norwegian (bokmål), a nasal or lateral consonant may in exceptional cases function as the nucleus of an unstressed syllable.

Syllabic nasals and laterals may have onset (e.g. the /d+/ in /'VE.d+n+%/) and the /k/ in /2'spe+:.kl%).sE). As the syllabic consonant derives from the elimination of an underlying non-stressed vowel (usually a schwa-type vowel) in the nucleus, the onset is included in order to keep the syllabic structure.

However, after <kk> (a double k) the <e> will be pronounced and there is no syllabification of the <l> unless followed by the suffix <else>:

Examples:

sykkel (Eng. bicycle)	ˈsY.kEɪ
nøkkel (Eng. key)	2ˈnE+.kEɪ
vekkelse (Eng. revival)	2ˈVE.kɪ%).sE

Syllabic <n> occurs after certain consonants, namely <d>, <l>, <n>, <r>, <s> and <t> and is in most cases a reduction of the enclitic articles of nouns (-en, -ene). Syllabic <l> occurs in the derivational suffixes <-else> and <sel> and when preceded by the consonants <d> and <t>.

Examples:

spaden (Eng: spade) (d+en)	2ˈspA:.dn%)
skolen (Eng: the school) (l+en)	2ˈsku:.ln%)
kommunen (Eng.: the county) (n+en)	kU.2ˈmu0:.nn%)
faren (Eng: the father) (r+en)	2ˈfA:.n+%)
dressene (Eng: the suits) (s+en)	2ˈdr6E.sn%).E
gutten (Eng: the boy) (t+en)	ˈgU0.tn%)
handel (Eng: business) (d+el)	ˈhAn.dɪ%)
tistel (Eng: thistle) (t+el)	ˈtɪ.stɪ%)
undersøkelsene (Eng: investigations) (-else)	2ˈU0.nE.ˈ2Se+:.kɪ%).sn%).E
hendelse (Eng: incident) (-else)	2ˈhE.nɪ%).sE
forkjølelse (Eng: a cold) (-else)	fO.ˈCe+:.ɪl%).sE
godtgjørelse (Eng: payment)	2ˈgOt.ˈ2je+:.ɪ+%).sE

Syllabic <n> and <l> become retroflex when preceded by the equivalent retroflexed variants of <d>, <l>, <n> and <t>:

Examples:

varsel (Eng: warning) (-sel)	ˈVA.Sɪ%) (not retroflex after <S>)
sporten (Eng: the sport)	ˈspO.t+n+%)

Words ending on <len> and <ren> (like *skolen* and *redaktøren*) are realized with deletion of the schwa in the last syllable, making the final nasal syllabic.

Examples:

skolen (Eng: the school)	2ˈsku:.ln%)
redaktøren (Eng: the publisher)	r6E.dAk.ˈte+:.n+%)

Also in words ending in <tn> the <n> becomes syllabic:

Examples:

vatn (Eng: water)	ˈVA.tn%)
botn (Eng: bottom)	ˈbO.tn%)

In many dialects, particularly in Eastern Norway, also retroflex pronunciation of <len> and <ren> is widespread: /sku:.n+%/ (*skolen*) and /lE&I.En+/
(*leiren*). This pronunciation is omitted here as it is considered vulgar.

The words ending on <elen> and <eren> (like *himmelen* and *sommeren*) are part of the same phenomenon as words ending on <len> and <ren>. In addition to deletion of the schwa in the last syllable, these words also have a syllable deletion - in other words the final nasal is not made syllabic.

Examples:

<i>Orthography</i>	<i>Singular definite</i>	<i>Plural definite</i>
himmelen (Eng: the sky, the heaven)	2'hI.mEln	2'him.ln%).E
sommeren (Eng: the summer)	2'sO.mEn+	2'sOm.r6E.nE

2.5 The postalveolar fricative (/S/)

2.5.1 A phoneme of its own

The postalveolar fricative is a phoneme of its own, and may as such occur in the onset of a syllable. There are certain phonotactic restrictions: In the chosen language variant the /S/ usually cannot occur in front of /V, m, n, p, t, k, pr, tr, kr/ without the word sounding exaggerated or vulgar. In standardized pronunciations the alveolar fricative /s/ occurs here.

Examples:

sjø (Eng: sea)	'Se+:
skjerm (Eng: screen)	'SEr6m
sky (Eng: cloud)	'Sy:
giro (Eng: giro)	'Si:.r6U
<i>but:</i> svær (Eng: huge)	'sV@:r6

2.5.2 <rs> → /S/

But /S/ can also occur as a phonological modification of the cluster <rs> (/rs/ → /S/, a process somewhat similar to the occurrence of retroflexes). Transcribing this instance of the /S/ in the onset clashes with the phonotactic rules of Norwegian, as does the transcription of retroflexes in onsets. However, if one considers the sandhi* process which leads to this variant of /S/ to be similar to retroflexation, then the /S/ as a result of the cluster <rs> goes in the onset in the same way as an /s/:

Examples:

forstå (Eng: understand)	fO.'Sto: (/S/ as result of <rs>)
forskrift (Eng: regulation)	2'fO.'2Skr6Ift (/S/ as result of <rs>)
forsvar (Eng: defence)	2'fO.'2SVA:r6 (/S/ as a result of <rs>)

2.5.3 <sl> → /SL/

The /S/ may, however, occur in front of /l/ without the connotations of vulgarity or exaggeration that arises if the sound is used in front of /V, m, n, p, t, k, pr, tr, kr/. In the recording of the spoken language variant, the morpheme-initial onset <sl> is pronounced /SL/. This pronunciation is becoming more widespread in the area of the chosen language variant,

* The term *sandhi* comes from sanskrit and means “joining”. It is used to describe phonological modifications of juxtaposed elements/segments.

which traditionally has the alveolar pronunciation, /sl/. This is most likely due to influences from Eastern Norwegian. Thus, we transcribe <sl> in morpheme-initial onsets as /Sl/. The <sl> in morpheme-internal or morpheme-final position is pronounced /sl/ in the chosen language variant, e.g.: *vesle* - 2'VE.sIE (Eng: small/tiny). The only exception to this is the pronunciation of the city name *Oslo*, which may be pronounced 2'U.SIU. This pronunciation is most likely due to the frequency of the city name (as the capital of Norway) and its pronunciation among Eastern Norwegian speakers in TV and radio.

Examples:

slag (Eng: speciem/en/stroke/beat)	'SlA:g
anslå (Eng: estimate)	'An.'2SlO:
Oslo (capital of Norway)	2'U.SIU (morpheme-internal /Sl/)

2.5.4 In separate morphemes and in compounds

Note that both the /s/ and the /l/ have to belong to the same morpheme in order to be pronounced /Sl/ morpheme-initially. In case the (sl) is a product of compounding, it is considered vulgar to pronounce the sequence as /Sl/ in the chosen language variant. Due to the MOP rule for compounds, these words are transcribed separately resulting in /s/ + /l/ anyway:

Examples:

Island (Eng: Iceland) *(Is + land)	'i:s. '2lAn
mislykket (Eng: unsuccessful)	2'ml.'2slY.kEt
huslig (Eng: domestic)	2'hu0:.sll
huslås (Eng: house lock) (hus + lås)	2'hu0:s.'2lo:s
idrettslag (Eng: sports club) (idrett+s+lag)	2'i:.dr6Ets.'2lA:g

* Cf. App. 2 – *Decompounding item 3.3*

2.6 The realization of the word-initial <tj>, <kj>, <ki> and <tsj>

In the chosen language variant, the <tj> may be realized as a palatal affricate /t&S/. The native speaker does so in almost all instances. However, the more standardized bokmål pronunciation is to pronounce this onset as /tj/ (not as an affricate, but as an alveolar plosive followed by a devoiced glide/semi-vowel) whenever the <tj> occurs in derivations and compounds of the word *tjene* (Eng: serve or Eng: earn). Otherwise, <tj> should be pronounced /C/ (palatal fricative). We follow this standardized variant.

Since /tj/ is a possible onset, all occurrences of <tj> across morpheme boundaries will be put in the onset (e.g. in affixations), but not across compound boundaries – cf. 1.1

Examples:

tjenester (Eng: services)	2'tje:.nE.stEr6
tjukk (Eng: fat)	'CU0k
utgift (Eng: expence)	2'u0:.'2tjIft (prefix: /u0:t/, stem:/jIft/)
<i>but:</i> matjord (Eng: topsoil) (mat + jord)	2'mA:t.'2ju:r6

The <kj> and <ki> are both pronounced as /C/ (palatal fricative), although it is possible in the chosen language variant to pronounce them as a palatal affricate, like the <tsj>. We use /C/.

Examples:

kjør (Eng: drive)	'Ce+:r6
kirke (Eng: church)	2'CIr6.kE

In loan words though, <ki> is pronounced /kI/ like in *kimono*, *kibbutz*

2.6.1 The affricate /t&S/

The <tsj> is not a possible Norwegian onset, nor does it denote a Norwegian phoneme. It occurs, however, in loan words from other languages. like *chat* (from English), *Tsjekkia* (from Czech), *Chile* (from Spanish), and in proper names and in derivations of such. This sound may in some spoken variants be pronounced as the fricatives /S/ or /C/, but the more standardized variant is the affricate /t&S/. Consequently, we transcribe the instances of this sound with the affricate.

Examples:

chat (from English: chat)	't&SEt
tsjekkisk (Eng: Czeck)	't&SE.kIsk
chilensk (Eng: Chilean)	t&SI.'le:nsk

2.7 The realization of the word-initial <gi> , <gy>, <gei> and <gøy>

Words of Old Norse origin starting with <gi>, <gy>, <gei> and <gøy> are pronounced as palatal approximants /jI/ , /jY/ , /jE&I/ and /jE+&Y/. Other words may be pronounced with a velar plosive: /gI/ , /gY/ , /gE&I/ and /gE+&Y/.

Words borrowed into Norwegian from other languages may be adapted to either of these pronunciations depending on what original language it came from, at what time it came and whether it came into use in oral or written language. There are no rules available for which words should be pronounced in what way as modern usage also may vary. For all existing entries in Vanvik, we refer to these.

Examples:

gift (Eng: poison)	'jIft
gitter (Eng: grate/bars)	'gI.tEr6

2.8 The realization of the word-initial <sk>

When <sk> in Norwegian words is followed by an <i> or a <y> in the same morpheme, the pronunciation will be /S/. In loan words with foreign pronunciation <sk> is pronounced /sk/

Examples:

ski (Eng: ski)	'Si:
skygge (Eng: shadow)	2'SY.gE
skin-head (English)	'skIn.'2hEd

3 VOWELS

3.1 /U/ or /O/?

The morphemes <om> and <opp> are transcribed /Om/ and /Op/. This is also the case whenever <om> and <opp> occur in polymorphemic words.

In western parts of Norway the <o> is more often pronounced /U/ than in the rest of the country. The native speaker also does so. However, we lean more to the Eastern Norwegian standard and will use an /O/ in marginal cases. The distribution of the two allophones cannot be found in written sources. Usage is obscure and conducted by geography, sex, age and social-economic belonging. In dubious cases we use *Vanvik*.

Examples:

gjennom (Eng: through)	'jE.nOm
gjenopprette (Eng: reestablish)	'jE.nO.'2pr6E.tE
konto (Eng: account)	'kOn.tU
kontor (Eng: office)	kUn.'tu:r6
kontroll (Eng: control)	kUn.'tr6Ol

3.2 Vowels in unstressed position

As a rule a vowel in unstressed position will be short and lax, cf. 3.3.3 below. In certain positions, however, we apply length markers to maintain the vowel quality of a stressed vowel. Thus all vowels may have 3 variants: the stressed, long and tense vowel ('e:), the unstressed, short and lax vowel (E) and the unstressed short and tense vowel (e:). The last one is used

- a) for vowels in unstressed syllables in compounds when the same syllable would have been assigned a primary or secondary stress in the simplex word. By compounds we mean grammatical compounds, not only the words which are decomposed according to our own conventions in our lexica
- b) in the suffixes -het, -skap, -bar, -vis, -tiv

Examples:

landbruk (Eng: agriculture)		2'lAn.'2br6u0:k
landbruksminister (Eng: Minister of agriculture)	landbruk+s+minister	2'lAn.br6u0:ks.ml.'2nI.stEr6
vennskap (Eng: friendship)		2'VEn.'2skA:p
vennskapsbånd (E:bond of friendship)	vennskap+s+bånd	2'VEn.skA:ps.'2bOn

Long vowels are not used in syllables before the main stress with a few exceptions, namely in derivations ending on a long and stressed /i:/ which receives the main stress of the word, when the word from which it originates had a long and stressed vowel:

Examples:

<i>Basic form</i>	<i>Transcription</i>	<i>Derivation</i>	<i>Transcription</i>
reder (Eng: ship owner)	2'r6e:.dEr6	rederi (company)	r6e:.dE.'r6i:

heler (Eng: receiver of stolen goods)	2'he:lEr6	heleri (the act of receiving)	he:lE.'r6i:
---------------------------------------	-----------	-------------------------------	-------------

3.3 The semi-vowel /j/

The semi-vowel /j/ is a phoneme and a possible onset in Norwegian. The same sound is audible also as a transitional sound between vowels when one of the vowels is a high front vowel like IPA /i, I, y, Y/. In these cases, the sound is not a phoneme and is thus not transcribed in a phonemic transcription. (Transitions between sounds are taken care of in the diphone library).

In more informal and rapid speech, a syllable reduction might take place when one of two adjacent vowels is a high front vowel. The high front vowel is reduced to the semi-vowel /j/ and might be transcribed as an onset before the second vowel. In these cases we include two transcriptions – one in which we transcribe a more formal pronunciation, thus keeping the full number of syllables, and one in which we transcribe a commonly reduced variant (duplicate).

Examples:

hjerne (Eng: brain)	2'j@:.n+E (/j/ as onset)
sier (Eng: say/says)	2'si:.Er6 (not 2'si:jEr6 – no transitional sound should be transcribed)
studie (Eng: study)	'stu0:.dI.E (not 'stu0:.dI.jE or 'stu0:.djE – no transitional sound should be transcribed and the number of syllables should be kept up.)
ingeniør (E: engineer)	In.SE.nI.'e+:r6 (formal)
	In.SE.'nje+:r6 (informal, rapid speech)

4. PROSODY

4.1 Stress

4.1.1 *Primary stress*

All isolated words are transcribed with primary stress.

Primary stress can manifest itself in three ways in Norwegian:

1. If the word is a monosyllable it receives primary stress only.

Example:

gå (Eng: go)	'go:	primary stress
--------------	------	----------------

2. If a polysyllabic word is stressed on the word-final syllable, it also receives primary stress only. In short, if there is no syllable following the primary stressed syllable, the word is only assigned primary stress.

Example:

nasjonal (Eng: national)	nA.SU.'nA:l	primary stress
--------------------------	-------------	----------------

Acronyms are an exception to this rule. They may receive secondary stress also in front of the main stress when pronounced letter by letter. This also applies to acronym compounds.

Examples:

PLO-lederen (Eng: the PLO-leader)	PLO+++lederen	‘2pe:.El.’u:.’2le:.dEn+
KFUM-leder (Eng: YMCA-leader)	KFUM+++leder	’2ko:.Ef.u0:.’Em.’2le:.dEr6

3. If the word is polysyllabic and receives stress on a word-initial or word-internal syllable, this syllable receives primary stress and toneme 1 or primary stress and toneme 2. In short, if there is a syllable following the primary stressed syllable, the word receives primary stress and a toneme.

Examples:

over (Eng: over)	‘o:.VEr6	primary stress and toneme 1
alene (Eng: alone)	A.2’le:.nE	primary stress and toneme 2

4.1.2 *Secondary stress*

Secondary stress does not receive tonemicity in Norwegian.

Secondary stress is linked to the notion of compounding. Generally speaking, we might say that in a compound, primary stress is assigned to the first constituent of the compound, whereas the other compound constituent receives secondary stress.

Examples:

telefonkatalog (E: teleph. directory)	telefon+katalog	tE.lE.’fu:n.kA.tA.’2lo:g	‘=prim.stress, ‘2=sec.stress
kirkebakke (Eng: church hill)	kirke+bakke	2’CIr6.kE.’2bA.kE	2’=pr.stress, ‘2=sec.stress

For acronyms, cf. 2.2.2 above.

When a word with both primary and secondary stress (which may itself be a compound) becomes part of a compound, the secondary stress is assigned to the last part of the compound.

Examples:

landbruk (Eng: agriculture)		2’lAn.’2bru0:k
landbruksminister (Eng. minister of agriculture)	landbruk+s+minister	2’lAn.br6u0:ks.mI.’2nI.stEr6

The notion of *compound* is not easily defined. Certain compounds have been lexicalized to the extent that no one thinks of the words as consisting of more than one constituent any longer, e.g. *hanske* (Eng: glove), etymologically derived from *hand* + *sko* (Eng: hand + shoe). The word *hanske* receives primary stress only – the latter constituent has lost all evidence of being part of a compound through stress weakening, but to say that this is a general rule for lexicalized compounds would be completely wrong. First of all, it is impossible to establish the border between lexicalized items and non-lexicalized items. Secondly, there are numerous other lexicalized compounds which receive both primary and secondary stress, e.g. *skohorn*,

kråskap, fuglebrett, fjernsyn. There is also the issue of compounding through affixation, where certain affixes are primary stressed, others are secondary stressed, and others yet again are never stressed.

Examples:

oppdage (Eng: discover)	‘Op.’2dA:.gE	prefix: opp-
kjærlighet (Eng: love)	2’C@:..l+I.’2he:t	suffix: -het
behandle (Eng: treat)	bE.’hAn.lE	prefix: be-

As mentioned previously, a compound word may consist of more than two constituents. Thus, some compounds may be assigned two or more secondary stressed syllables. Stress may be assigned in a hierarchical order, with primary stress on top and secondary stress of varying degrees in a falling order. The word *vaktmesterassistent* (six syllables, three constituents) might therefore be perceived as having primary stress on the first syllable (*vakt-*), secondary stress of the first degree on the last syllable (*-stent*), and secondary stress of the second degree on the second syllable (*-me-*). In our transcription we will not distinguish between varying degrees of secondary stress, but merely transcribe secondary stress on the syllable with most stress next to the primary stressed syllable.

A compound is thus assigned one primary and one secondary stress.

Example:

vaktmesterassistent (Eng: caretaker assistant)	vaktmester+assistent	‘VAkt.mE.stEr6.A.sI.’2stEnt
motforestillinger (Eng: objections)		2’mu:t.fo:.r6E.’2stI.lInK.Er6
råvarepristilbud (Eng: price estimate of raw materials)	rå+vare+pris+tilbud	2’r6o:.VA:.r6E.pr6i:s.tIl.’2bu0:d

In Norwegian compounds there may be a shifting of the stress in the second (and/or following) constituent (-s). This is to avoid a "clash" between two stressed syllables and maintain a balance within the compounded word. When as a separate word the constituent would have had main stress on the first syllable (2’utstyr), it may receive secondary stress on the following syllable when it is second part of a compound (-ut’2styr). For this to happen, the main stress of the first constituent must be on its last syllable (kon’tor-).

Examples:

utstyr (Eng: equipment)		2’u0:t.’2sty:r6	primary stress on <ut>
kontorutstyr (Eng: office equipment)	kontor+utstyr	kUn.’tu:r6.u0:t.’2sty:r6	<ut> is unstressed, while <styr> receives the secondary stress

4.1.3 Stress and tonemicity in foreign words

Foreign words pronounced in a foreign way are only assigned stress, not tonemicity. If the pronunciation is norwegianized we apply toneme indication.

Example:

<i>Orthography</i>	<i>Code</i>	<i>Transcription</i>	<i>Code</i>
whiteboard (from English)	eng.WWW	ˈVA&It.ˈ2bo:d+	eng.WWW
Sierra Leone	for.XYZ	sI.ˈE.r6A ˈIE.2ˈo:.nE	nor.NOR

4.2 Length

4.2.1 *Length in primary stressed syllables*

Norwegian distinguishes between two kinds of length in the rhyme structure of syllables: V(long)C(short) and V(short)C(long). The length differences are manifested phonetically as durational differences and in some cases spectral differences between phonemically long and phonemically short vowels.

The vowel length is thought to be primary, while consonant duration adapts to vowel duration in order to maintain the overall rhyme duration. In light of this, length is marked only on phonemically long vowels and these are transcribed with the length marker (:).

Example:

ansvarlig (Eng: responsible)	An.ˈsVA:.l+I
------------------------------	--------------

4.2.2 *Length in secondary stressed syllables*

The durational difference between segments in primary, secondary and unstressed syllables is taken care of through the use of different symbols for primary and secondary stress and no symbol for unstressed syllables.

Phonemically long vowels are transcribed with length markers also in syllables with secondary stress. The suffixes –het, -bar, -vis, -skap, -messig, -aktig -som and -tiv are assigned secondary stress and lengthmarkers in simplex words.

Examples:

fuglebur (Eng: birdcage)	2ˈfu0:.lE.ˈ2bu0:r6
ansvarlighet (Eng: responsibility)	An.ˈsVA:.l+I.ˈ2he:t
ekteskap (Eng: marriage)	2ˈEk.tE.ˈ2skA:p

4.2.3 *Length in unstressed syllables*

Length markers are also applied to vowels in unstressed position in certain suffixes and in compounds when the vowel would have received stress in the simplex word, cf. 1.11 and 3.2 above.

Example:

ansvarlighetsfølelse (Eng: feeling of responsibility)	ansvarlighet+s+følelse	An.ˈsVA:.l+I.he:ts.ˈ2fe+:.lI%).sE
---	------------------------	-----------------------------------

lårbensbrudd (E: fracture of femur)	lår+ben+s+brudd	2'lo:r6.be:ns.'2br6U0d
-------------------------------------	-----------------	------------------------

The unstressed syllable has an underlyingly long vowel.

4.3 Tonemicity

In most Norwegian dialects we have 2 tonemes. In some dialects this phonological opposition is non-existent, and speakers of these dialects are usually unable to recognize the difference when listening to speakers who have them. It is impossible to give exhaustive rules for the occurrence of the two respectively, although a number of rules can be derived from morphology. But in many verb/noun homographs, the toneme is significant, cf. 1.4 above.

In some compounds both tonemes may be accepted as normal. The choice is arbitrary. For the sake of consistency we have decided to use toneme 2 in these cases.

Example:

sammenlagt (Eng: put together)	sammen+lagt	2'sA.mEn.'2lAgt
offentlig (Eng: public)		2'O.fEnt.lI
gjensidig (Eng: mutual)		2'jEn.'2si:.dI

5. OTHER SPECIFIC TRANSCRIPTION CONVENTIONS

5.1 <veg> and <høg>

According to Norwegian orthography the noun 'veg' (road) and the adjective 'høg' (high) have alternative spelling 'vei' and 'høy'. By most people speaking *bokmål*, they will be pronounced with the diphthongs /E&I/ and /E+&Y/. We have therefore included alternative transcriptions.

Examples:

landeveg (Eng.: country road)	2'lA.nE.'2VE&I (first alternative)
	2'lA.nE.'2Ve:g (second alternative)
høgskole (Eng: college)	2'hE+&Y.'2sku:.lE (first alternative)
	2'he+:g.2sku:.lE (second alternative)

5.2 Foreign words ending on <-or>

Words from Latin ending on <or> may have a shifting of the stress in plural. It is more and more common, however, to accept a pronunciation without this shifting. We have included both, having the nativized pronunciation as number one.

Examples:

traktorer (Eng: tractors)	'tr6Ak.tU.r6Er6 (first alternative)
	tr6Ak.'tu:.r6Er6 (second alternative)
professorer (Eng: professors)	pr6U.'fE.sU.r6Er6 (first alternative)

	pr6U.fE.'su:.r6Er6 (second alternative)
--	---

5.3 Words in ‘nynorsk’

As these conventions are outlining pronunciation for the language variant ‘bokmål’, ‘nynorsk’ words which may occur in the lists are transcribed the way they would be pronounced by a speaker of ‘bokmål’. These words are coded “nno.NOR”

Example:

<i>Orthography</i>	<i>Transcription</i>	<i>Code</i>
gladaste (Eng: happiest)	2'glA:.dA.stE	nno.NOR

5.4 Words in ‘riksmål’

Archaic forms of bokmål called “riksmål” are not part of the official Norwegian spelling standard but may be found in private usage. Such words may be included in our transcription lexicon, for instance <efter>, <sne>. They receive the code for bokmål, nor.NOR

5.5 Multiple word utterances

Multiple word utterances have an underscore between word boundaries in orthography and should also have so in decompounding and in transcription. In the decompounding field they have as a rule a +-sign before and after the underscore, but there are exceptions.

Words with an underscore in the transcription are transcribed in the same way as each of them would have been transcribed as individual entries. An exception is function words where reduced forms are applied.

Multiple word utterances receive a *phrase accent*. In the transcription the word with the main stress is assigned an extra stress indicator (“) in front of that word in addition to the stress/toneme indication on the syllable carrying the main stress. The other words are assigned a secondary stress or a main stress depending on the pronunciation of the phrase as a whole.

Cf. Appendix 6 – Multiple word phrases, Appendix, 4 – Names, item 8 and Appendix 2 – Decompounding, 3.3

5.6 <y> in unstressed position in names

Names ending in an unstressed <y> may be pronounced with an /y/ or and /i/. For such names we only transcribe the alternative /Y/ (cf. Appendix 4 – Names):

Examples:

Willy	‘VI.IY (also pronounced /‘VI.II/)
Johnny	‘jO.nY (also pronounced /‘jO.nI/)

5.7. Assimilation vs. phonemic transcription

5.7.1 Devoicing

An unvoiced consonant (p, t, k, f, s) following a voiced consonant (b, d, g, v) will have a devoicing influence on the voiced consonant. The transcriptions will not mirror this process, however, even though the assimilation is sometimes complete:

Examples:

<i>Orthography</i>	<i>Transcription</i>	<i>Pronunciation</i>
AIDS	ˈE&Ids	ˈE&Its
til_havs (to sea)	ˈ2tIl ˈhAVs	ˈ2tIl ˈhAfs
lagt	ˈlAgt	ˈlAkt

5.7.2 Transitional changes due to compounding

In compounds with a long vowel and a voiced consonant in the last syllable of the first constituent, this vowel will be reduced in spoken language and the consonant be devoiced when the syllable is followed by a linking-s. Still, compounds are transcribed as if they were two or more separate words, keeping up the vowel and consonant qualities of each word. The transcription does not mirror the factual vowel reduction and consonant devoicing that may take place.

Examples:

<i>Orthography</i>	<i>Transcr.</i>	<i>Compounding</i>	<i>Transcription</i>	<i>Pronunciation</i>
dag (Eng: day)	ˈdA:g	dag+s+aktuell	ˈdA:gs.Ak.tU0.ˈ2E1	ˈdAks.Ak.tU0.ˈ2E1
tid (Eng: time)	ˈti:d	tid+s+bestemt	ˈti:ds.bE.ˈ2stEmt	ˈtIts.bE.ˈ2stEmt
krig (Eng: war)	ˈkri:g	krig+s+helt	ˈkr6i:gs.ˈ2hElt	ˈkr6Iks.ˈ2hElt
li:v (Eng: life)	ˈli:v	liv+s+viktig	ˈli:vs.ˈ2Vlk.tI	ˈIifs.ˈ2vIk.tI

Also a final <n> after a short vowel in the last syllable of the first constituent followed by a <g> pronounced /g/ in the next constituent will make up /nK.g/ in pronunciation as a consequence of the transition. If the word is decomposed however, it will not be reflected in the transcription.

Example:

<i>Orthography</i>	<i>Transcr.</i>	<i>Compounding</i>	<i>Transcription</i>	<i>Real pronunciation</i>
vann (E:water)	ˈVAn	vann+glass	2ˈVAn.ˈ2glAs	2ˈVAnK.ˈ2glAs

5.8 Phonological variants

In cases where alternative transcriptions can be generated by rules we only transcribe one version of the word. Whenever two or more alternatives are equally frequent in use and there is no such general pattern, we transcribe two (or more) variants. Cf. 5.1.

Example:

<i>Orthography</i>	<i>Alternative I</i>	<i>Alternative II</i>
berekne (Eng: estimate)	bE.'r6E&I.nE	bE.'r6E.knE
meteorology (E: meteorologist)	mE.tE.U.r6U.'lo:g	mE.trU.'lo:g
amerikaner (E: American)	A.mE.r6I.'kA:.nEr6	Am.r6I.'kA:.nEr6

5.9 Acronyms and abbreviations

Acronyms are sometimes pronounced letter by letter (NRK) and sometimes as an ordinary word (NATO). The transcription should mirror the way the acronym is used in spoken language. For special rules, cf. 4.1.2.

Acronyms should as far as possible have an expansion, but this should not be transcribed.

For abbreviations only the expanded form is transcribed:

<i>Orthography</i>	<i>Expansion</i>	<i>Transcription</i>
f.eks	for_eksempel	'2fOr6_'Ek.'sEm.pEl

6. Diacritics and suprasegmentals

6.1 Diacritics and suprasegmentals currently in use

The use of diacritics is limited to one since the transcription is to be as phonemic as possible. In addition many diacritic characteristics will be taken care of in the diphone library and rules for synthesis.

The only diacritic in use is: Syllabic mark for nasals and laterals. The syllabic alveolar nasal and lateral are listed as allophones in the PhonTable v1.08 for Norwegian, but for practical purposes they are listed as phonemes in the phoneme table for PhonEd.

Suprasegmentals in use are:

- primary stress/toneme 1 (´)
- toneme 2 (2´)
- secondary stress (˘2)
- length mark (:)
- syllable boundary (.)
- phrase accent (˘)
- underscore (˘)

APPENDICES

Appendix til Transcription Conventions for Norwegian v1.5 rev.1.05

Appendix 1 -**v. 1.3**

Dette vedlegget er utarbeidet av språkteknikerne Berit Veka og Rut Krogset ved Nst, Voss.

Tabell 1: Mulige opptakter i norsk

1 KONSONANT			2 KONSONANTER			3 KONSONANTER		
L&H+	Ortografisk	Transkribert	L&H+	Ortografisk	Transkribert	L&H+	Ortografisk	Transkribert
p	pil	'pi:l	pr6	pris	'pr6i:s	spr6	språk	'spr6o:k
b	båt	'bo:t	pl	plagg	'plAg	spl	spleise	2'spIE&l.sE
t	tog	'to:g	pj	pjølter	'pjOl.tEr6	spj	spjelke	2'spjEl.kE
d	dag	'dA:g	br6	bra	'br6A:	str6	strev	'str6e:V
k	kar	'kA:r6	bl	blå	'blo:	stj	stjele	2'stje:.lE
g	går	'go:r6	bj	bjørn	'bje+:n+	skV	skvise	2'skVi:.sE
m	mann	'mAn	tr6	trapp	'tr6Ap	skr6	skrik	'skr6i:k
n	natt	'nAt	tV	tvil	'tVi:l	skl	skli	'skli:
f	fin	'fi:n	tj	tjene	2'tje:..nE			
V	våt	'Vo:t	dV	dvele	2'dVe:..lE			
r6	ring	'r6lnK	dj	djerv	'djEr6V			
s	si	'si:	dr6	dra	'dr6A:			
S[1]	sjal	'SA:l	kn	knuse	2'knu0:..sE			
	skjørt	'SE+t+	kV	kvikk	'kVlk			
	sky	'Sy:	kr6	kropp	'kr6Op			
	giro	'Si:..r6U	kl	klem	'klEm			
C[2]	kjøpe	2'Ce+:..pE	gn	gnu	'gnu0:			
	tjukk	'CU0k	gr6	grå	'gr6o:			
	kino	'Ci:..nU	gl	glad	'glA:			
h	har	'hA:r6	mj	mjød	'mje+:d			
l	liv	'li:V	nj[4]	Njål/linjal	'njo:l			
j[3]	jage	2'jA:..gE	fn	fnis	'fni:s			
	hjelp	2'jEl.pE	fr6	frekk	'fr6Ek			
	gjøre	2'je+:..r6E	fl	flår	'flo:r6			
	gyng	2'jYnK.E	fj	fjell	'fjEl			
	gi	'ji:	Vr6	vrang	'Vr6AnK			
			sp	spor	'spu:r6			
		st	stol	'stu:l				
		sk	skal	'skAl				
		sm	små	'smo:				
		sn	snill	'snll				
		sV	svak	'sVA:k				
		Sl	slå	'Slo:				
[1] L&H+ [S] kan ha både <sj>, <skj> and <sk> som ortografi. <gi> blir uttalt [Si] ved foreldet skrivemåte av <giraff>.								
<gi> og <gy> kan ha både [gi] og [gy] som ortografi i egennavn og i lånnord.								
[2] L&H+ [C] kan ha både <kj>, <tj> and <k> i ortografi. Unntak: <tjeneste>.								
[3] L&H+ [j] kan også ha <hj>, <g> og <gj> som ortografi.								
[4] Kombinasjonen [nj] i begynnelsen av et ord forekommer bare i egennavn og stedsnavn.								

I lånnord kan andre opptakter forekomme, for eksempel <sf> i sfære, <ps> i psykolog osv.									
Tabell 2: Spesielle tilpasninger - gjelder inni ord									
1 KONSONANT			2 KONSONANTER			3 KONSONANTER			
L&H+	Ortografisk	Transkribert	L&H+	Ortografisk	Transkribert	L&H+	Ortografisk	Transkribert	
t+	artig	2'A.t+l	Sp	forspill	2'fO.'2SpIl	Skr6	forskrift	2'fO.'2Skr6lft	
d+	ferdig	2'fE.d+l	Sm	spørsmål	'spE+.'2Smo:l	Spr6	forsprang	2'fO.Spr6AnK	
n+	barna	'bA:.n+A	Sn	forsnakke	fO.'SnA.kE	Str6	forstrekke	fO.'Str6E.kE	
l+	ærlig	2'@:.l+l	SV	forsvar	2'fO.'2SVA:r6				
S	persisk	'pE.SlSk	sl	vesle	2'VE.slE				
			St	første	2'fE+.StE				
t&S	chat	't&SEt	Sk	forsker	2'fO.SkEr6				
Også retrofleksene [t+], [d+], [n+] og [l+] kan stå som opptakt, men bare foran vokaler og inni ord.									
Sammensmeltingen <rs>, [S], kan stå som opptakt inni ord, foran vokaler eller foran konsonantene									
[p], [m],[n], [V], [l], [k], [kr6], [tr6] og [pr6].									
Affrikaten /t&S/ forekommer bare i lånnord, egennavn og avledninger av disse.									

Appendix til Transcription Conventions v1.5, rev.1.05
 hnø 16.08.00

Appendix 2 - Deling av ord

Versjon 3.06

1. Definisjoner

- 1.1 Et sammensatt ord (kompositum) kan defineres som et ord sammensatt av minst to frie morfemer, dvs. to ord med individuell leksikalsk betydning.
- 1.2 *Morfem*: En minste betydningsbærende enhet i et språk.
Bøyningsmorfem: Bøyningsendelse, for eksempel –ere (komparativ) i ordet rundere, -t (nøytrum) i ordet rundt. Bøyningsmorfemet kommer etter alle andre morfemer i et ord og kalles derfor ofte for bøyningsuffix.
Avledningsmorfem: -ing i ordet skiving, u- i ordet utydelig, billed- i ordet billedskjønn. Avledningsmorfemer kan stå både foran et annet morfem (avledningsprefiks) og etter et annet morfem (avledningssuffix).

Rotmorfem: Morfemer som normalt kan fungere som selvstendige ord.
Stamme: Den delen av et ord som går igjen i alle bøyningsformer. Ordene barn, barna og barnehage har alle stammen barn.
Fugemorfem: Disse morfemene opptrer i overgangen mellom de viktigste delene av visse sammensatte ord, for eksempel barnehage, seierssrus, lammekotelett.
Grammatisk morfem: Avledningsmorfem, fugemorfem og bøyingsmorfem er grammatiske morfemer og tilhører den lukkede klassen av grammatiske ord (jfr. for eksempel preposisjoner).
Semantisk morfem: Rotmorfemene kalles også semantiske morfemer da de i de fleste tilfeller tilhører de åpne ordklassene substantiv, verb og adjektiv.
Fritt morfem: Disse kan fungere som stammer av selvstendige ord. Gjelder først og fremst rotmorfemene.
Bundet morfem: Disse forekommer oftest som deler av ord og består først og fremst av grammatiske morfemer.
Restmorfem: Morfemer som ikke fungerer som stamme og heller ikke forekommer mer enn i enkelte tilfeller sammen med et annet morfem. Jfr. for eksempel kirs- i kirsebær, tytte- i tyttebær.
Pos (Part of speech) : Ordklasse
Decpos (Decompounded part of speech): Ordklasse på hver enkelt del av et dekomponert ord.

2. Generelle regler for deling

Sammensatte ord skal deles når hvert ledd har samme eller nærliggende betydning til den det får i sammensetningen. Sammensetningsrensens angis med et plusstegn (+) mellom ordleddene *skoreparatør* kan deles i *sko + reparatør*, *fjernkontroll* i *fjern + kontroll*, *sykehusvesen* i *syk+e+hus + vesen*.

Når et ord/en frase blir delt, skal hver del tildeles en ordklasse (decpos).

Om deling av navn: se under pkt. 3.3 og Appendix 4 – Names

Merk: *skomaker* skal ikke deles da *maker* ikke kan stå alene. (Det er en derivasjon = avledning, jfr. *blikkenslager*, *våpendrager*) Det samme gjelder *tyttebær*, *hovedbruk* da verken *tytte* eller *hoved* er selvstendige morfemer.

* To eller flere frie substantiv:
 hus+eier: hus+eier+forening; skole+år

*	<u>Adjektiv + substantiv:</u> blå+farge; blind+gate syk+e+hus (<i>grunnform av adj.</i> <i>når det er mulig</i>) grønt+areal	<u>Substantiv + adjektiv:</u> mote+bevisst; miljø+vennlig
---	--	--

*	<u>Verb + substantiv:</u> gå+stol; spaser+sko bade+plass (<i>se under</i>) reise+feber	<u>Substantiv + verb:</u> kniv+stikke
---	---	--

Der første ledd tilsynelatende kan være både et substantiv med linkemorfem og et verb, velger vi verb i infinitiv (jfr. bade+plass, ikke bad+e+plass).

* Adverb + Substantiv
straks+tiltak

a) Adjektiv + Adjektiv**
ferdig+gjort

b) Adjektiv + verb**
ferdig+gjort

**Mrk: 'gjort' er en partisipp av verbet 'å gjøre'. Partisipper kan ofte brukes som adjektiv. For å tilfredsstille konvensjonen om at hele ordets PoS skal være den samme som siste ledds decomp, velger vi i eksempel a) å kalle partisippet for et adjektiv. Slike ord må dupliseres da de også kan være verb (eksempel b)).

Egennavn (propria) skal betraktes på samme måte som andre ord og deles (se nedenfor og egen konvensjon.)

3. Spesielle regler for deling

3.1 Affikser og forledd

Prefiks og suffiks skal ikke deles fra.

Vi deler følgelig ikke opp ord som begynner på f. eks. be-, gjen-, hen-, ad-/at-, for-, på-, eller som ender på f. eks. -bar og -aktig. Det samme gjelder affikser som bio-, anti-, -messig, -dom osv.

Eksempler på ord som ikke skal deles:

Prefiks	Suffiks	Forledd
beundre		
forbehold		
	skjønnhet	
	kjørbar	
	melkaktig	
		elektromontør
		økokrim
		mikrobiolog

Ord som slutter på -løs skal deles. (tannløs → tann+løs)

Sammensetninger med preposisjoner deles vanligvis ikke. Men ord der preposisjonen har beholdt sin konkrete betydning, deles. Det samme gjelder sammensetninger som kjennetegnes ved at de kan brukes løst sammensatt med samme betydning:

* preposisjon + verb:
over+bygg (konkret betydn.)
av+kle (men ikke 'avgi')
av + venne

preposisjon + substantiv:
over+bygg (men ikke 'overtid')
over+trekk+s+bukse

- * Akronym + substantiv:
CD+-+spiller; NST+-+ansatt

3.2 Sammensetninger med fugemorfem

Fugemorfem er morfem som fungerer som bindeledd mellom to sammensetningsledd. De hyppigst forekommende linkemorfemer i norsk er <s> og <e>. Men også <n> og <er>, <ar>, <a> og <me> kan forekomme:

s:	mor+s+rollen
e:	barn+e+hage; mus+e+felle
n:	rose+n+kål
er:	berlin+er+bolle; rett+er+gang
ar:	laug+ar+dam
a:	ferd+a+folk
me:	lam+me+kotelett

Fugemorfem markers med plusstegn (+) foran og etter.

Bindestrek anses som fugemorfem i de tilfeller der den har som funksjon å binde sammen (se under).

3.3 Egennavn (propria) deles slik: (Jfr. Appendiks 4 – Navn)

- * Når personnavn, geografiske navn, navn på bygninger og konstruksjoner er sammensatt av to eller flere deler som hver for seg er selvstendige, meningsbærende ord, skal de deles. Når første ledd kan være et fellesnavn, skal første bokstav være liten i dekomponeringsfeltet. Cf. App. 4 – Names, item 4.

Ortografi	Deling	Decpos	Pos	Transkripsjon
Drammensveien	drammen+s+veien	noun+punc+noun	ng	'dr6A.mEns.'2VE&I.En
Storelund	stor+ e+lund	adj+link+noun	npl	2'stu:.r6E.'2IU0n
Lilleberg	lille+berg	adj+noun	npl	2'II.IE.'2bEr6g
Hardangerfjorden	Hardanger+fjorden	ngo+noun	ngo	hA.'d+AnK.Er6.'2fju:.n+%)
Lars-Ole	Lars+-+Ole	npfm+punc+npf m	npf m	'2IA:S.'2u:.IE

Når navnet (f.eks. et gatenavn) består av flere ord, skilles ordene med “underscore” (_) i ortografien og transkripsjonen. De dekomponeres ikke.

Ortografi	Transkripsjon
Ole_Irgens_vei	'2u:.IE '''Ir6.gEns '2VE&I
Klaus_Hansens_vei	'2kIE&U0s '''2'hAn.sn%)s '2VE&I
Ture_Nermans_vei	'2tu0:.r6E '''ne:r6.mAns '2VE&I
Åsane_Torg	2'o:.sA.nE '''tOr6g/

Slike navn får også en **fraseaksent** som tildeles det ordet som får hovedtrykket i helheten. Det markeres med anførselstegn (double quote) (") foran ordet slik at det får dobbel markering ("" eller "2") for eksempel /'2A.nE_''kA.'r6i:.nE/. Vanlig trykk og tonelag markeres som ellers. De øvrige ordene i navnet behandles som frittstående ord, men tildeles det trykk og tonelag de får i helheten (frasen), se eksempel over og under:

Ortografi	Transkripsjon
Lyder_Sagens_gate	'2Lyder ""Sagens '2gate
Ole_Irgens_vei	'2Ole ""Irgens '2vei
Klaus_Hansens_vei	'2Klaus ""Hansens '2vei
Ture_Nermans_vei	'2Ture ""Nermans '2vei
Åsane_Torg	2'Åsane ""Torg
Mexico_by	'Mexico ""by
Sierra_Leone	Si'2erra ""Le2'one eller: Si'erra ""Le2'one

- * Når første ledd har en form som bare forekommer i sammensetninger, skal ordet merkes med et utropstegn (!) umiddelbart foran ordet i dekomponeringsfeltet, og deles:

Lofotveggen → !Lofot + veggen

Dersom sammensetningsformen kommer som annet enn 1. ledd, skal ordet ikke deles og heller ikke markeres med utropstegn (boogie-woogie).

- * Fellesnavn som er sammensatt av et egennavn (proprium) og et fellesnavn (appellativ) deles:

Willoch-regjeringen → Willoch +-+ regjeringen

willochregjeringen → !willoch + regjeringen

grønlandshval → !grønland+s+hval

Dersom 1. ledd i et sammensatt egennavn er et fellesnavn, skriver vi det med liten forbokstav i dekomponeringen (jfr. 3.3, 1. avsnitt):

Torgalmenningen → torg + almenningen

Merk: Hvert av leddene får PoS ,noun', mens PoS for hele ordet følger ortografifeltet, i tilfellet over ,ngs' (street name). Dette er et unntak fra regelen om at PoS for hele ordet skal være det samme som decpos for siste ledd.

Egennavn deles ikke når

- ortografien er ulik den i tilsvarende fellesnavn: Nygaard, Stendahl.
- navnet slutter på *-sen/-son*
- uttalen av navnet som helhet fraviker uttalen av hver del: <Selbu> /'sEl.'2bu0/ vs. <sel> /se:l/

3.4.a Ord med bindestrek

- skal markeres+-+.....

Bindestrek anses som fugemorfem i de tilfeller der den har som funksjon å binde sammen. Den markeres da i dekomponeringen med plusstegn foran og etter (FN-sjef → FN +-+ sjef). Når bindestreken har andre funksjoner, som for eksempel betydningen fra-til, skal den ikke skilles ut (Bergen-Voss).

Den skal heller ikke skilles ut i sammensetninger med forstavelser som ellers ikke deles fra (bio-ingeniør, øko-melk).

I ord med bindestrek der to eller flere av leddene er gjensidig avhengig av hverandre for å gi mening, deles de ikke:

Eksempel:

<i>Ortografi:</i>	<i>Deling:</i>	<i>Ordklasser (decpos):</i>	<i>Transkripsjon:</i>
ping-pong-ball	ping-pong+-+ball	noun + noun	'pInK.pOnK.'2bAl

Bindestrek markeres ikke i transkripsjonen.

3.4.b Flerordsuttrykk med bindestrek

Når et flerordsuttrykk er satt sammen med bindestrek med et annen ord, betraktes flerordsuttrykket som en helhet som utgjør den ene delen av sammensetningen:

Eksempel:

<i>Ortografi:</i>	<i>Deling:</i>	<i>Ordklasser (decpos):</i>	<i>Transkripsjon:</i>
a_la_carte-servering	a_la_carte+-+servering	noun + noun	A.lA.'kAt+.sEr6'2ve:r6InK

4. Andre regler

4.1 Egne sammensetningsformer

Når et sammensatt ord består av en egen sammensetningsform + 2. ledd (jfr. ovenfor) skal ordet merkes med utropstegn (!) umiddelbart foran ordet i dekomponeringsfeltet og deles.

!billed+skjønn; !lofot+fiske

4.2 Feilstavete ord

Dersom et ord er feilstavet, sjekk først i "alle transkriberte" om ordet allerede finnes med rett ortografi. I så fall merkes ordet med "garbage". Hvis det ikke finnes, skal det rettes der det står og deles.

4.3 Dobbelkonsonanter

Når ett ledd slutter på en dobbel konsonant og neste ledd begynner på samme konsonant, skal ordet markeres med utropstegn (!) umiddelbart foran ordet i dekomponeringsfeltet og den doble konsonanten forenkles:

busstasjon = buss + stasjon → !bus + stasjon

5. Ordklasse: Siste ledd i et kompositum skal ha samme ordklasse som hele ordet. Et unntak her er sammensatte egennavn der siste ledd kan være et fellesnavn, jfr. pkt. 3.3 over.

Appendix til Transcription Conventions for Norwegian v.1.5, rev.1.05
10.07.00 – hnø

Appendix 3 – Utenlandske ord **Versjon 1.2**

Transkripsjonskonvensjoner for utenlandske ord

Transkripsjon av utenlandske ord må tilpasses det foneminventaret som finnes i L&H+ for norsk. Vi kan ikke trekke inn symboler fra andre språk (moderspråkene) og ikke finne opp nye symboler eller kombinasjoner av de eksisterende.

Det kan imidlertid være nødvendig å fravike konvensjonene for mulige opptakter, jfr. navn som Sri Lanka. <sr> er ingen mulig opptakt i norsk, men må her aksepteres.

Dersom ordet kan uttales på flere måter, skal vi transkribere 2 varianter:

1. én så lik den vanligste måten å uttale ordet på norsk som mulig: <juice> /ju0:s/
2. én så lik uttalen på originalspråket som mulig: <juice> /dju0:s/

Utenlandske ord får *trykk*, men vanligvis ikke *tonelag*.

Svært få andre språk har tonelag. Ord fra lite kjente språk får vanligvis en uttale som er tilpasset norsk uttale og kan dermed få tonelag.

Eksempler:

Sierra Leone /sI.'E.r6A_'IE.2'o:.nE/ (fornorsket)
Besserwisser /'bE.sEr6.'2VI.sEr6/

Eksempler på "vanskelige" fonemer:

(Andre språk enn engelsk er markert i parentes)

- De stemte og ustemte frikativene i engelsk (som i <them> og <think>) transkriberes som sine tilsvarende klusiler:

them = /dEm/

think = /tInKk/

- Nasale vokaler (fransk) transkriberes som den tilsvarende orale vokalen fulgt av /nK/:

blanc (fr.) = /'blAnK/

Champs Elyssés (fr.) = /'2SanKs_''E.II.'se:/

- Stemte affrikater transkriberes som sin motsvarende stemte klusil+frikativ:
George = /'djo:t&S/

- Ustemt affrikat er et lovlig fonem, jfr. Appendix 1, "Possible onsets":

chat = /'t&SEt/

chicken = /'t&SI.kEn/

- Diftonger som ikke finnes i det norske inventaret transkriberes med "nærmeste" eksisterende diftong:

town = /'tE&U0n/

Blaupunkt (tysk) = /'bIE&U0.'2pUnKkt/

- Labio-velar approksimant /w/ transkriberes som /V/

word = /'Ve+:d+/

Washington = /'VO.SInK.tn%/

Anvendte koder:

Engelsk	eng.WWW
Fransk	fre.WWW
Tysk	ger.WWW
Andre språk (unknown)	for.XYZ

Appendix to Transcription Conventions for Norwegian v.1.5, rev. 1.05
16.08.00 – Gisle Andersen/Helga Nygaard Øverland

Appendix 4 – Names ***version 1.4***

Person names, geographic names

(For transcriptions based on Onomastica, see version 1.3)

GENERAL INFORMATION

Garbage marking

Names which should be removed are marked with a *g* in the comment field.

Person names

Entries that contain (or consist solely of) abbreviated names (*Chr, Aasm, Th, Andr*, etc.) or one or more initials (*Per Ø, Kari H F*), and those that contain a first name and a part of the surname should be marked *garbage*, e.g.:

Pål von
Herman de
Maria van der
Jan van de
Gunnhild de la
Robert le

The same applies to entries consisting of the abbreviations *jr, sr, dy, de*. In other words, entries like *Per dy* and *Laurits jr* should be marked as garbage.

Misspellings

Obvious misspellings should not be corrected, but simply be marked as garbage.

Street names

Some street names consist of an abbreviated first name and a surname: Chr. Mickelsens gate, C. Sundts gate. If the full form of the abbreviation is known, it should be transcribed accordingly. (/’2kr6l.stl.An.’2’ml.kl%).sn%)s_’2gA:.tE/ and /’2se:’”sU0nts_’2gA:.tE/) If the abbreviation is unknown, mark the word ** in front of the orthography and leave it, but do not mark it ‘garbage’.

Abbreviated forms of “gate” and “vei” should be written out in full orthography and transcription.

POS-tagging

The personal names should be tagged according to gender.

Double names

Generally speaking, double names may be pronounced with a primary stress on each element or on one element only. We only want one pronunciation represented in the lexicon, namely the one that carries one primary stress only. In most cases, the primary stress must be assigned to the last element, especially in names that follow the very common monosyllabic + disyllabic pattern, eg Jan-ERIK. In some cases, the primary stress may be assigned to the first element, for instance if the compound consists of two monosyllabic names, eg. MAY-Britt. In addition, the least phonologically salient element should be assigned a secondary stress (cf. Transcription conventions below).

In double names a secondary stress may be applied in front of the main stress (cf. acronyms and multiple word phrases).

Example: Stene+-+Johansen is transcribed /’2ste:.nE.jU.’hAn.sn%)/

Double first names

There should not be duplicates of the type *Jan Erik / Jan-Erik*. (Duplicates will be generated automatically at a later stage)

Double surnames

Double surnames spelled with a hyphen should be treated as compounds and follow the same pattern of pronunciation as double first names. <Stene-Johansen> should thus be decomposed into Stene+-+Johansen and transcribed /'2ste:.nE.jU.'hAn.sn%/).

Phonological variants

More than other words names may be pronounced in different ways, cf. below.

Transcription conventions

1. For transcription of names, use the general “Transcription Conventions for Norwegian v.1.5, rev. 1.03” including Appendixes 1-3 (March 2000) as guidelines. However, spelling and pronunciation of proper nouns are not subject to the same rules as for common nouns. The names can therefore not always be transcribed phonemically. (Ex: <Grieg> → /gr6Ig/, <Hiim> → /hi:m/)
2. For each entry four variants are possible (4 fields).
- 3a) For alternative transcriptions: Concentrate on variants which are as general and widespread as possible, always taking the most common as first choice. Local dialect variants should be avoided.

Examples:	Karin	1) /2'kA:.r6In
		2) /kA.'r6i:n
	John	1) /'jOn/
		2) /'jUn/
		3) /'ju:n/
	Karen	1) /2'kA:.n+%)/
		2) /2'kA:.r6En/

3b) Names ending in an unstressed <y> may be pronounced with an /y/ or and /i/. For such names we only transcribe the alternative /Y/ (cf. Trans. Conv. p. 20, item 5.6):

Example:

Willy	'vI.IY (also pronounced /'vI.II/)
Johnny	'jO.nY (also pronounced /'jO.nI/)

4. Decomposing

Names (proper nouns) may be decomposed according to Appendix 2, v.3.05 – Decomposing, 3.3.

Compounded names are treated as other compounds with primary and secondary stresses and syllable boundaries. When the first constituent may be a common noun, it is spelled with a lower case first letter in the decompounding: <Bakkeby> = bakke+by.

It is often hard to tell the origin of proper names. Whenever a name may be split into two or more constituents which separately have their own meaning and may be assigned a PoS, they should be decompounded even though this meaning is not necessarily ethymologically correct. (Example: Rønneberg → rønne (common noun, indefinite form, singular, m+f) +berg, Holbygda → Hol (geographic noun) + bygda (common noun, definite form, singular, feminine)). Avoid absurdities, though. (cf. Appendix 2 – Decompounding)

Additionally, each constituent must have fairly the same pronunciation as simplex words as they have in the compound. The name ‘Selbu’ /’sEl.ʔ2bu0:/ is not split as ‘sel’ is differently pronounced as a simplex word: /’se:l/

Names are not decompounded when one of the constituents has a deviating orthography from the corresponding common noun. (Nygaard, Rodahl, not: Ny+gaard, Ro+dahl)

Also, names are not decompounded when other constituents than the first one is a proper noun. (Annelise, not Anne+Lise, Sandsveen, not: Sand+Sveen, Nygaard, not Ny+Gaard even though Lise as well as Sveen and Gaard exist as surnames.)

Multiple-word names (Øvre_Flatåsveien) are not decompounded, and there is no +-sign between the words.

Remember: Concatenating the components, i.e. leaving out the +-sign, should always leave you with the original compound word.

5. Double names spelled with a hyphen are transcribed as simplex words and follow the maximal onset principle (MOP). Example: Reiss-Andersen /’2r6E&I.2’sA.nE.Sn%)/
6. The orthography should reflect the orthography as it is used in postal addresses. Thus ‘vei’ or ‘veg’ is a separate word, while ‘veien’ is compounded: “Hardangerveien” vs. “Ture Nermans vei”. The same applies to ‘gate’ and ‘gaten’, ‘allé’ and ‘alléen’.
As a reference we use Postadressebok for Norge, 21st edition, 1999.
7. A space in the orthographic form has to be reflected by an underscore (_) in the transcription.

8. Multiple word names

General information

Multiple word names have an underscore (_) in the orthographic string and should also have an underscore in the decompounding field as well as the transcription field, between word boundaries. When there is a hyphen between two of the words in a multiple word name, these two words are treated as a simplex word in the transcription, which implies use of MOP.

Example: Gunnar_Reiss-Andersens_gate /2'gU0.nAr6_'r6E&I.2'sA.nE.Sn%)s_'2gA:.tE/

Stress levels in multiple word names

All words in a multiple word name shall have a stress marking – either primary or secondary stress – and a sentence accent (phrase accent).

All multiple word phrases, and only those, can have three types of stress markings:

- sentence accent (the double quote) (“)
- primary stress (‘ for toneme one and 2’ for toneme two)
- secondary stress (‘2)

Symbols for sentence accent, primary stress and secondary stress shall be placed according to the following rules:

- 8.1 The symbol for sentence accent (“) can only occur once in a multiple word phrase. The sentence accent falls on the name that has the main accent of the whole utterance (e.g. Irgens in Ole_Irgens_vei. The symbol for sentence accent is the double quote (“) and is placed in front of the name that attracts the main accent of the whole utterance. All other stress symbols are placed according to general rules and indicate the syllable that attracts the stress of that particular word.
- 8.2 Primary stress is placed on names or syllables that are pronounced with a primary stress.
- 8.3 Secondary stress is placed on names or syllables that are pronounced with a secondary stress or that are unstressed in the utterance. If such a name originally has toneme 2, they lose their tonemicity when assigned a secondary stress mark in unstressed position: <2'Ole>, but <'2Ole_'Irgens_'2vei>
- 8.4.1 Function words like prepositions are assigned a secondary stress and are transcribed with their reduced forms.
- 8.4.2 For multiple word phrases consisting of words with equal stress distribution, i.e. where there is no clear peak on any word, there still has to be a sentence accent on one of the words. We will by default place the sentence accent on the last word in the phrase in those cases.

Examples:

Names

Ole_Irgens_vei	'2u:.lE_'r6.gEns_'2VE&I	
Universitetet_i_Oslo	U0.nI.VE.SI.'te:.tE_'2I_'2'U.SIU	(red.prep.)
Storstuen_på_Verket	2'stu:. '2Stu0:.En_'2pO_'VEr6.kE	(red.prep.)
Kristiansand_sør	kr6I.stI.An.'sAn_'se+:r6	
Søndre_Villavei	2'sE+n.dr6E_'VI.lA.'2VE&I	
Åsane_Torg	2'o:.sA.nE_'tOr6g	
Nedre_Bellevue	2'ne:.dr6E_'bE.lE.'Vy:	

Appendix 5 - Coding of transcriptions ***version 1.01***

Antallet felt som benyttes i et transkripsjonsarbeid vil variere fra oppdrag til oppdrag. I referanseleksikonet er det satt opp 40 felt, jfr. Appendix 7. hvorav 10 felt er satt av til språkspesifikk informasjon. Følgende oversikt gir opplysninger om koder og utfylling av noen av feltene.

1. Coding of PoS

1.1) Common parts of speech:

adj
adverb
conjunction
determiner
interjection
noun
numeral
preposition
pronoun
unknown
verb

1.2) Proper nouns:

n	noun common
nc	company/brand name
ngc	country name
ngi	city name
ngo	other geographic name
ngs	street name
no	organization name
np	person name (unspecified)
npf	first name
npff	female first name
npfm	male first name
npfn	sex-neutral name
npl	last name

All POS-codes exist as macros in the file Ldb on "Main"/No transcription/Lasse

Some words may be denoted more than one PoS. In these cases the entire entry is duplicated.

Examples:

<retter>	verb	<malte>	adjective
<retter>	noun	<malte>	verb

Some words may have various pronunciations while maintaining the same PoS. They will have 2 alternative transcriptions: The most widely applied pronunciation will always be no.1

Examples:

Finansbransjen 1) fl.'nAnKs.'2br6AnK.Sn%)

2) fl.'nAns.2'br6An.Sn%)

1.3 PoS on constituents in compounds (cf. App. 2)

Hyphen is tagged as "punc"

Linking elements are tagged as "link"

Examples:

sommer+ferie	noun+noun
Willoch+--+regjeringen	npl+punc+noun
selskap+s+damen	noun+link+noun

2. Coding of Gender

All nouns are assigned a gender code. Nouns which may be assigned more than one gender have multiple tags:

f = female
m = male
n = neutral

blindgate	blind+gate	adj+noun	fm
-----------	------------	----------	----

Adjectives in neuter form are also tagged for gender:

vintergrønt	vinter+grønt	noun+adj	n
-------------	--------------	----------	---

In compounds only the last constituent is assigned gender:

vintersola	vinter+sola	noun+noun	f
------------	-------------	-----------	---

Language coding:

All entries in the transcription lexicon are assigned language codes for orthography and for transcription. If a word is written in Norwegian orthography, it is marked as such. If it is written with the orthography of the language it originates from, it is assigned the code of that country. It consists of 3 lower case letters and 3 upper case letters.

1. For Norwegian bokmål it is *nor.NOR*, for nynorsk it is *nno.NOR*.
2. Foreign words are assigned the code for the language in question plus WWW (cf. below).

<i>Orthography</i>	<i>Code</i>	<i>Transcription</i>	<i>Code</i>	<i>Transcription</i>	<i>Code</i>
juice	eng.WWW	'ju0:s	nor.NOR	'dju0:s	eng.WWW
besserwisser	ger.WWW	'bE.sEr6..'2vI.sEr6	nor.NOR	-	-

3. If the origin is unknown to the lexicographer, the word should be assigned *for.XYZ*
4. If a word has two entries in orthography – one with a nativized and one with the original orthography – the first is assigned the code *nor.NOR* and the latter the code of the originating country (e.d. *eng.WWW*).

5. If a word has two entries in pronunciation – one with a nativized and one with an approximately original pronunciation – the first is assigned the code nor.NOR and the latter the code for the originating country.
6. If a word is a mixture of native and foreign, use the code for the foreign language.

Example:

<i>Orthography</i>	<i>Code</i>	<i>Transcription</i>	<i>Code</i>
juicemugge	nor.NOR	'dju0:s.'2mU0.gE	eng.WWW

7. If a word is a derivation of a foreign name, use the code of the foreign language

Example:

<i>Entry word</i>	<i>Language code</i>
Gorbachovisk	for.XYZ (Russian)

Codes for foreign languages:

Norsk	nor.NOR (bokmål)
Norsk	nno.NOR (nynorsk)
English	eng.WWW
French	fre.WWW
German	ger.WWW
Foreign (unknown)	for.XYZ

4. Ranking of transcriptions

The transcription conventions give guidelines for ranking of transcriptions whenever there is more than one alternative.

- | | |
|--------------|-------------|
| 1) fl.'nAnKs | 2'br6AnK.SE |
| 2) fl.'nAns | 2'br6An.SE |

5. Coding of Status and Certainty

The Status entry will always be STANDARD, and will always be part of the entry of a word.

Certainty: Should always be denoted.

0: should not be used. The word should rather be marked 'garbage'.

1: Transcription is done according to the knowledge of the language technician. No reference has been used.

2: The word is transcribed according to a reference lexicon.

6. The Comment field

The "Comments" field is enclosed by brackets ()

Garbage: Garbage words are tagged '(g)'

- Words with a wrong spelling which already exist with correct spelling in the transcription lexicon (search in the file "Alle transkriberte")
- Nonsense words

Acronyms and abbreviations:

Acronym	NST	(akr (Nordisk Språkteknologi))	'2En.Es.'te:
Abbreviation	UiO	(abr (Universitetet I Oslo)	U0.nI.VE.SI.'te:.tE_'2I_'"2U.SIU

7. Miscellaneous

- In case of uncertainty: Mark the word ****xxxxxx** to make it easily retrievable
- A manual for UltraEdit gives instructions how to
 - 1) make direct keys
 - 2) use macros
 - 3) search efficiently

.....

Appendix to Transcription Conventions for Norwegian v. 1.5., rev. 1.05
14 August 2000 – Catharina Kylander/Helga Nygaard Øverland

Appendix 6 – Multiple word phrases version 1.2

Instruction for transcription of names and other multiple word phrases **For names: cf. App. 4 – Names**

General information

Multiple word phrases have an underscore (_) in the orthographic string and should also have an underscore in the transcription, between word boundaries. In the decompounding field underscores should as a rule have a +-sign before and after. The PoS (decpos) should be 'punc'.

Example:

for_eksempel	for+_eksempel (prep. + punc + noun)	'2fOr6_'Ek.'sEm.pEl
--------------	-------------------------------------	---------------------

However, when one of the words in a multiple word phrase cannot occur on its own, but only as part of the phrase, the +-sign should not be inserted in the decompounding field and there is no decpos. Only the whole phrase will receive a PoS (ditt_og_datt, de_luxe, à_la_carte).

Example:

à_la_carte	à_la_carte (noun)	'2A_'2lA_'kAt+
------------	-------------------	----------------

If a multiple word phrase is part of a compound written with a hyphen, it should be decomposed and transcribed like other compounds:

Example:

à_la_carte-meny	à_la_carte+++meny (noun + noun)	'2A_ '2lA_ '''kAt+.mE.'2ny:
-----------------	---------------------------------	-----------------------------

NB! Words with a hyphen (-) in the orthography are NOT multiple word phrases and are thus considered as simplex or compound words and transcribed according to applicable conventions. When considered a compound word, MOP should not be applied, while it should be applied when the word is considered a simplex word.

Examples:

<i>Orthography:</i>	<i>Decompounding:</i>	<i>Transcription:</i>
stand-in	stand-in * (as simplex)	'st@:n.'2dIn (MOP applied)
cha-cha-cha	cha-cha-cha (as simplex)	't&SA.t&SA.t&SA
Willoch-regjeringen	Willoch+++regjeringen (as compound.)	'VI.lOk.r6E.'2je:.r6InK.En (MOP not applied)

* *foreign words are not decomposed*

Stress levels in multiple word phrases

Since the words in a multiple word phrase are separated by underscore, they are all considered as separate words, and therefore **all words in a multiple word phrase shall have a stress marking** – either primary or secondary – and a phrase accent (sentence accent).

All multiple word phrases, and only those, can have three types of stress markings:

- sentence accent (“)
- primary stress (‘ for toneme one and 2’ for toneme two)
- secondary stress (‘2)

The symbol for sentence accent (“) can only occur once in a multiple word phrase. It shall be placed in front of the word that attracts the main stress of the whole utterance. The word accents are then placed according to general rules and indicate the syllable that attracts the stress of that particular word.

Symbols for sentence accent, primary stress and secondary stress shall be placed according to the following rules:

Distribution of stress markings

1. The sentence accent falls on the word that has the main accent of the whole “utterance” (e.g. Irgens in Ole Irgens vei. The symbol for sentence accent is the double quote (“) and is placed in front of the word that attracts the main accent of the whole utterance. All other stress symbols are placed according to general rules.
2. Primary stress is placed on words that are pronounced with a primary stress. (Ex. b)

3. Secondary stress is placed on words (or syllables in words assigned a main stress) that are pronounced with a secondary stress or that are unstressed in the utterance.
4. The default transcription will give a clear pronunciation of all words in the word group. An exception is function words such as prepositions, which shall be assigned a secondary stress and be transcribed with their reduced forms. (cf. ex. b and c)
5. For multiple word phrases consisting of words with equal stress distribution, i.e. where there is no clear peak on any word, there still has to be a sentence accent on one of the words. We will by default place the sentence accent on the last word in the phrase in those cases. (Ex. d, e and f.)

Examples:

Names

a) Ole_Irgens_vei	'2u:.1E_'''Ir6.gEns_'2VE&I
b) Universitetet_i_Oslo	U0.nI.VE.SI.'te:.tE_'2I_'2'U.SIU
c) Storstuen_på_Verket	2'stu:.'2Stu0:.En_'2pO_'''VER6.kE
d) Kristiansand_sør	kr6I.stI.An.'sAn_'''se+:r6
e) Søndre_Villavei	2'sE+n.dr6E_'''VI.lA.'2VE&I
f) Åsane_Torg	2'o:..sA.nE_'''tOr6g

Abbreviations:

bl.a. (blant annet)	'2blAnt_'2'A.n%)t
f.eks (for eksempel)	'2fOr6_'Ek.'sEm.pEl

Lexicalized phrases

ditt_og_datt	'dIt_'2O_'''dAt
til_og_fra	'tIl_'2O_'''fr6A:

Reflexive verbs

bite fra seg	'2bi:.tE_'''fr6A:_'sE&I
vaske_seg	'2'VA.skE_'2sE&I

Verbs with particles

vaske_ned	'2VA.skE_'''ne:d
falle_sammen	'2fA.lE_'''sA.mEn

Other types

de_luxe	'2dE_'''1U0ks
---------	---------------

Appendix til Transcription Conventions for Norwegian v.1.5 rev. 1.05

hnø

Appendix 7 – Feltene i transkripsjonsleksikonet

Versjon 1.0

Hvilke felt som blir benyttet, vil variere fra ordre til ordre.

Feltene i transkripsjonsleksikonet er basert på manualen for boksene i PhonEd: Manual PhonEd v.1.70, rev. 1.01

Følgende felt gjelder for

Referanseleksikonet (100 K)

0. opprinnelig ortografi
1. korrigert ortografi
2. dekomponering
3. PoS for hele ordet
4. PoS for hver del i det sammensatte ordet
5. kjønn
6. språk, ortografisk
7. 'garbage'
8. akronym
9. forkortelse
10. ekspansjon av akronym/forkortelse
11. transkripsjon alt. 1
12. 'certainty' transkr.alt.1
13. status for transkr.alt.1
14. språk transkr. alt. 1
15. transkripsjon alt. 2
16. 'certainty' transkr. alt 2
17. status for transkr. alt 2
18. språk transkr. alt. 2
19. kommentarfelt
20. serienummer i opprinnelig transkripsjonsjobb
#21 – 30 er satt av til språkspesifikke felter
31. ordsett-spesifikk intern frekvensordning
32. review-felt
33. ordsett-spesifikk ordinndeling
34. ordsett-tilhørighet (for original ortografi, "finnes i")
35. transkriberingssett ("transkribert i")
40. unik id for post i vårt leksikon

.....

Appendix til Transcription Conventions for Norwegian v.1.5 rev. 1.05
hnø

Appendix 8 - Norske koda

Versjon 1.03

15.08.00

Alle enkle konsonanter kan stå i kodaposisjon, unntatt /h/, /j/ og /C/. Også affrikaten /t&S/ og velar nasal /nK/ kan stå som koda og regnes da henholdsvis som 3 og 2 tegn. Parseren teller tegn og ikke lyder.

To konsonanter (tegn) (bokstavene i første kolonne skal følge dem i andre kolonne)

s - p,b,t,d,k,g,m,n,f,V,l,nK
t - p,k,g,m,n,f,S,l,s, V,
l - g, nK, V
p - m, r6, l, s
d - V, l, nK, d, n
n - V, nK, m
g - r6, l,
k - r6, S, l, nK, s
m - r6, l,
f - l
V - l
S - t+, n+, l+, d+, n
b - m
j - l
f - m
l+ - n+
t+ - n+

Tre konsonanter (tegn) (bokstavene i første kolonne skal følge dem i andre kolonne)

sk - l, t, p, m, n
ts - l, s, m, k, f, n
s - mp, r6g, Vn, r6p, lp, lk, lg, ld, lm, lf, lV, r6V, r6m, gl, Sk, r6k, nKk, sk, mn, Vl, gd, ln, nd, pt, mf,
lj
t - r6k, Vn, ms, r6m, r6p, gs, ks, lg, lk, nK, r6V, r6f, sk, n+S, ns, fs, vs
d - mn, Vn

Fire konsonanter (tegn) (bokstaven i første kolonne skal følge dem i andre kolonne)

s - mnd, Vnd, nst, r6ft
t - lsk, nKk

Fem konsonanter (tegn) (bokstaven i første kolonne skal følge dem i andre kolonne)

s - nKkt

Eksempler på ord med ovenstående kodesammensetninger:

To konsonanter (tegn)

s - slaps, klubbs, sats, aids, saks, fags, hems, grunns, tufs, havs, hals, tangs,

t - slapt, takt, sagt, klamt, sundt, saft, børst, kaldt, hest, lavt

l - hagl, hangl, avl

p - hamp, varp, hjelp, asp

d - kløvd, hold, sprengd, legd, stipend

n - navn, sagn, hamn

g - sorg, salg

k - sterk, harsk, hulk, bank, rask

m - varm, halm

f - alf

V - alv

S - farts, Karls, lunsj, atferds, verns

b - bomb

l - detalj

f - trumf

l+ - Arnljot

t+ - internt

Tre konsonater (tegn)

sk - elsk, skotsk, løpsk, lumsk, svensk

ts - pulst, busts, amts, slekts, lofts, pasients

s - klumps, sorgs, navns, varps, hjelps, hulks, salgs, skylds, halms, gulfs, alvs, vervs,

sverms, hagls, karsks, snorks, tanks, rusks, hamns, avls, legds, himmelens, stipends, Egypts, triumfs, detaljs

t - mørkt, jevnt, varmt, skarpt, fulgt, mulkt, djervt, verft, raskt, Ernst, minst

st - sammenkomst, flittigst, oppvekst, fangst, kunst, øvst, nifst

d - nemnd, nevnd

Fire konsonanter (tegn)

s - nemnds, nevnds, kunsts, verfts

t - falskt, tenkt

Fem konsonanter (tegn)

s - punkts

