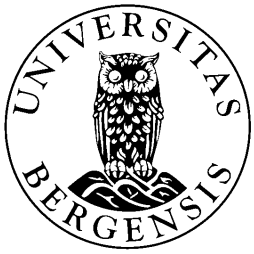

Project ref. no.	*LE3-4239*
Project title	*SCARRIE*
Deliverable status	*Restricted*
Contractual date of delivery	*July 31, 1997*
Actual date of delivery	*February 27, 1998 (modified March 10, 1999)*
Deliverable number	*D-3.3.1*
Deliverable title	***Specification of Tagset***
Type	*SP*
Status & version	*Final, version 7*
Number of pages	*16*
WP contributing to the deliverable	*WP 3*
WP/Task responsible	*UIB-HIT, Allégaten 27, 5020 Bergen, Norway*
Author(s)	*Victoria Rosén, Koenraad de Smedt*
EC Project Officer	*Antonio Sanfilippo*
Keywords	*morphosyntactic categories, morphosyntactic features, dictionary entries, written norms, gender systems*
Abstract	*The tagset designating categories and features to be used in the Norwegian dictionary is specified in this document.*

DEL 3.3.1 Tagset

This document specifies the categories and features necessary for the SCARRIE lexicon. When the category or feature corresponds to one used in NorKompLeks, the NKL category or feature is given in parentheses after the desired SCARRIE category or feature. When there is no corresponding category or feature in NKL, an asterisk appears.

Dictionary information in a CORRie-compatible 'intermediate' dictionary has six fields:

1. word form

2. frequency: this is not dealt with in this document

3. 'style' information: this is mainly used for (a) marking words which only occur in multi-word expressions and (b) handling written norms

4. morphosyntactic features, including word class, relevant for the compound grammar

5. replacements, dependent on field 3

6. syntactic features, including word class, relevant for the sentence grammar

This document addresses the coding of fields 3, 4, 5 and 6. The categories and features for fields 4 and 6 are discussed in section 1. The format in this document is, unless noted otherwise, the one used in field 6. The formats are similar, but there are some important differences. In field 4, categories and features are separated from each other by commas. In field 6 they are separated by underscores.

Features which are relevant only for the sentence grammar are to be listed only in field 6, not in field 4. This holds e.g. for the verb complementation and noun complementation features discussed below. In addition, the following features do not occur in field 4, only field 6: acc, active, actpass, aux, comp, count, countuncount, def, defindef, dem, e, expl, f, fm, fn, indic, infpres, int, m, main, mf, mfn, mn, n, nogender, p1, p123, p2, p3, passive, pastpart, pers, poss, prespart, rec, refl, sup, uncount.

Very few features which are relevant only to compounding are listed in field 4 but not field 6. This concerns the features 'part' discussed under Verbs below and the feature 'one' used for certain determiners.

1. Categories and features

The categories for SCARRIE are, with the exception of proper nouns and the special categories 'endepart' and 'infimp' (cf. the section Verbs below), the same ones as in NKL, but with different names.

The following categories have no features:

Aa (inf_merke)
Adv (adv)
Conj (conj)
Interj (interj)
P (prep)
RelPro
Subj (subj)

endepart
infimp

The categories which have features are the following:

Adj (adj)
Det (det)
N (subst)
PN
PossDet
Pref
Pro (pron)
V (verb)

As a general rule, each SCARRIE lexical entry within a single category must have the same number of features. There are only two exceptions to this. The first exception will be cases in which there is a single word form which represents two or more morphosyntactic words and where there is more than one feature in which these morphosyntactic words differ. An example is given for adjectives. The second exception is cases in which the same word form corresponds to morphosyntactic words belonging to different categories. An example is given for verbs.

When a feature is unspecified for certain word forms in a category, a dummy feature may be used. This dummy feature should be the string 'no' followed by an abbreviation of the morphosyntactic category involved, for example 'nogender' for unspecified gender.

N: Nouns

N covers nouns except proper nouns (PN).

All entries in the category N have features for gender, number, definiteness, countability, and complement. Some nouns also have a written norm feature.

Gender:
- m (m)
- f (mf, f)
- fm (mf, f)
- n (n)
- nogender (*)
- mf (mf) en art, ei art; en krokodille, ei krokodille
- mfn (mfn) en vis, ei vis, et vis
- mn (mn) en drops, et drops; en kilo, et kilo
- fn (mfn, fn) ei gardin, et gardin

The gender feature is assigned according to a language norm which has three genders. Word forms of feminine nouns are normally coded 'f', except the 'masculine' definite singular, which is coded as 'fm'. For example, *boken* is coded 'fm', all other forms of the noun (*bok, boka, bøker, bøkene*) are coded as 'f'. (This is a change with respect to version 1 of this document.)

Nouns which are 'm' or 'n' are assigned 'mn' (cf. Norsk Referansegrammatik, p. 153)

Nouns which are 'm' or 'f' are assigned 'mf' (ibid.)

Nouns which are 'f' or 'n' are assigned 'fn' (ibid.)

Nouns which are 'm', 'f' or 'n' are assigned 'mfn' (ibid.)

The feature 'nogender' is applied to words which never show gender agreement, e.g. *Pascal, aino, aktiv*, etc.

Number: sg (sg)
pl (pl)
sgpl (*)

Number is 'sgpl' if the wordform can be either singular or plural, e.g. *hus.*

Definiteness: def (best)
indef (ubest)
defindef (be_ub)

The feature 'defindef' is used if the word form can be either definite or indefinite, e.g. *stormsentra.*

Countability: count (*)
uncount (*)
countuncount (*)

The feature 'count' is assigned to countable nouns. The feature 'uncount' is assigned to mass nouns and other uncountables, e.g. *sand.* The feature 'countuncount' is assigned to nouns which can be both, e.g. *melk, en melk.*

Complement: takespl (*)
takesmass (*)
takesplmass (*)
nocomp (*)

The feature 'takespl' is assigned to quantifier nouns which can take a complement and require it to be plural, e.g. *en rekke personer, en kurv epler.* The feature 'takesmass' is assigned to quantifier nouns which require a mass noun if they take a complement, e.g. *et stykke ost.* The feature 'takesplmass' is assigned to quantifier nouns which require a mass noun or plural if they take a complement, e.g. *et kilo sukker, et kilo bønner.* The feature 'nocomp' is assigned to nouns which cannot function as quantifiers. (cf. NRG 3.3.1.1, p. 236 ff.)

Summing up, the features for nouns should be given in the following order: gender, number, definiteness, countability, complement.

Examples:

bok	N_f_sg_indef_count_nocomp
boka	N_f_sg_def_count_nocomp
boken	N_fm_sg_def_count_nocomp

PN: Proper Nouns

Number: sg (sg)
pl (pl)
sgpl (*)

NKL does not seem to contain any proper nouns. Examples like the following must probably be considered coding errors:

nkl_ff('Basedows sjukdom',[subst,sg,propr,4688,'Basedows sjukdom']).

Adj: Adjectives

As a rule, entries in the category Adj have features for gender, degree, definiteness, and number. Some entries however, may have a single feature 'e' rather than features for definiteness and number (see below).

Gender: m (m)

f (f)
n (n)
mf (mf)
mfn (mfn)

The feature mf is used for word forms that may agree with either a masculine or a feminine noun, e.g. *bakteriell.* The feature mfn is used for word forms that may agree with nouns of any gender, e.g. *balinesisk.*

Degree: pos (pos)
comp (komp)
sup (sup)

Definiteness: def (best)
indef (ubest)
defindef (ubest_best)

The feature 'defindef' is used for an adjective form which is either 'def' or 'indef', e.g. *bra.*

Number: sg (sg)
pl (pl)
sgpl (sg_pl)

The feature 'sgpl' is used for an adjective form which is either 'sg' or 'pl'. The combination 'defindef', 'sgpl' is used for an adjective which has only one form for all combinations of definiteness and number, e.g. *lunknere.*

Definiteness
and Number: e (*)

The feature 'e' is applied when an adjective is 'pl def', 'pl indef' or 'sg def'. This is the case for adjectives declined with the '-e' suffix, e.g. *gode,* and some others, e.g. *elskede.* In this case the adjective has only three features.

Summing up, the features for adjectives should be given in the following order: gender, degree, number, definiteness.

Examples:

lunken	Adj_mf_pos_indef_sg
lunkent	Adj_n_pos_indef_sg
lunkne	Adj_mfn_pos_e
lunknere	Adj_mfn_comp_defindef_sgpl
lunknest	Adj_mfn_sup_indef_sg
lunkneste	Adj_mfn_sup_def_sgpl

V: Verbs

Entries in the category V normally have features for tense/form, mood, voice, verb type, and valency.

Tense/Form: pres (pres)
pret (pret)
prespart (pres_part)
pastpart (perf_part)
inf (infinitiv)
infpres (*)

The passive infinitives do not seem to be present in NKL. The code 'infpres' should be used whenever the infinitive is expressed by the same word form as the present, e.g. *snakkes, gjøres,* etc. (combined with the passive feature for Voice).

Mood: indic (indikativ)
imp (imperativ)

Voice: active (aktiv)
passive (passiv)

Verb Type: main (hovedverb)
aux (*)

The feature 'aux' must be applied to auxiliaries (this results in separate entries when a wordform is both 'aux' and 'main').

Valency: dummysubj (nullv, nullv1)
intrans (intrans1–intrans4, trans11, trans12, trans13, part4, part5, refl13, adv3–adv5)
trans (trans1, trans5–trans10, trans14, ditrans4–ditrans6, part1, part3, refl14, adv6, adv7, adv11)
ditrans (ditrans1, adv10)
ncomp (predik1, predik4)
acomp (predik2)
objncomp (predik3, predik5)
objacomp (predik6, predik7)
scomp (trans2, trans15, trans16, trans18)
objscomp (ditrans2)
infcomp (trans3, trans17, trans19)
objinfcomp (ditrans3)
ref (refl1–refl5, refl9–refl12, part2, adv2, adv8, adv9)
transref (refl6)
refscomp (refl7)
refinfcomp (refl8)
pres (pres1–pres3)
accinf (trans4)

When a verb can have several valency features, they should be combined into a single feature. For instance, if the verb *spise* has the features 'intrans', 'trans' and 'objacomp' (*han spiser, han spiser middag, han spiser seg mett*), these should be combined to a single feature 'intrans\objacomp\trans'. When creating such complex features, the simple features should be ordered alphabetically and separated by backslashes.

The correspondences for valency indicated in the above list are for the active forms of the verb. Since passive reduces the valency of verbs, the passive verb forms will have different valency features than the active verb forms. The following table indicates the general correspondences expected between active and passive valencies.

ACTIVE	PASSIVE
dummysubj	no passive counterpart
intrans	dummysubj
trans	intrans

ditrans	trans
objncomp	ncomp
objacomp	acomp
scomp	intrans
objscomp	trans, scomp
infcomp	trans
accinf	infcomp
ncomp	no passive counterpart
acomp	no passive counterpart
objinfcomp	no passive counterpart
ref	no passive counterpart
transref	no passive counterpart
refscomp	no passive counterpart
refinfcomp	no passive counterpart
pres	no passive counterpart

For certain verb forms there are special categories which are used instead of category and all features except for valency. These categories, like the 'e' feature for adjectives mentioned above, make it possible to reduce the size of the word form list by replacing multiple entries of the same form. These categories are given in the following table. (The special features 'apart', 'etpart', 'dtpart', 'partadj' and 'pretadj' proposed in the earlier versions of this document are no longer used, since they did not allow for correct style replacements.)

SPECIAL CATEGORY	REPLACES
endepart	Adj_mfn_pos_defindef_sgpl V_prespart_indic_active_main
infimp	V_inf_indic_active_main V_pres_imp_active_main

In the following, the conjugation codes used in BMO are referred to. In all four verbal conjugations (v1–v4), the present participle is identical in form to the adjective; for these the code 'endepart' is used. In v4, the infinitive and the imperative are identical; here the code 'infimp' is used.

Summing up, the features for verbs should be given in the following order: tense/form, mood, voice, verb type, and valency. When one of the special categories is used, it should be separated from the valency feature by an underscore.

Examples:

spark	V_pres_imp_active_main_trans
sparka	Adj_mfn_pos_indef_sg
sparka	Adj_mfn_pos_e
sparka	V_pret_indic_active_main_trans
sparka	V_pastpart_indic_active_main_trans
sparka	V_pastpart_indic_passive_main_intrans
sparke	V_inf_indic_active_main_trans
sparkede	Adj_mfn_pos_e
sparkende	endepart_trans
sparker	V_pres_indic_active_main_trans
sparkes	V_infpres_indic_passive_main_intrans
sparket	Adj_mfn_pos_indef_sg
sparket	V_pret_indic_active_main_trans

sparket	V_pastpart_indic_active_main_trans
sparket	V_pastpart_indic_passive_main_intrans
sparkete	Adj_mfn_pos_e

A special feature is in addition necessary for the compound analyzer. The feature 'part' should be found in the fourth column for all pastparticiples, whether they belong to the category verb or adjective.

Examples of fourth column field:

spark	V,pres,imp
sparka	Adj,pos,indef,sg,part
sparka	Adj,pos
sparka	V,pret
sparka	V,part
sparka	V,part
sparke	V,inf
sparkede	Adj,pos,e
sparkende	endepart
sparker	V,pres
sparkes	V
sparket	Adj,pos,indef,sg,part
sparket	V,pret
sparket	V,part
sparket	V,part
sparkete	Adj,pos,e

Pro: Pronouns

Pronouns have features for gender, case, number, person, definiteness, and pronoun type.

Gender:	m (m)
	f (f)
	n (n)
	mf (mf)
	mfn (mfn)

The feature 'mf' should be applied to pronouns that can agree with words of masculine or feminine gender, e.g. *den.* The feature 'mfn' should be applied to pronouns that can agree with words of any gender, e.g. *seg.*

Case:	nom (nominativ)
	acc (akkusativ)
	nomacc (*)

The feature 'nomacc' should be applied whenever a wordform is either nominative or accusative, e.g. *han.*

Number:	sg (sg)
	pl (pl)
	sgpl (sg_pl)

The feature 'sgpl' should be applied whenever a word form is either singular or plural, e.g. *seg.*

Person:	p1 (p1)
	p2 (p2)

p3 (p3)

Definiteness: def (best)
indef (ubest)

Pronoun Type: pers (pers)
rec (res)
refl (refl)
int (int)

Summing up, the features for pronouns should be given in the following order: gender, case, number, person, definiteness, and pronoun type.

Examples:

den	Pro_mf_nomacc_sg_p3_def_pers
seg	Pro_mfn_acc_sgpl_p3_def_refl

Det: Determiners

Determiners have features for gender, number, definiteness and determiner type.

Gender: m (m)
f (f)
n (n)
mf (*)

The feature 'mf' is applied to determiners which can agree with words of either masculine or feminine gender, e.g. *den.*

Number: sg (sg)
pl (pl)
sgpl (sg_pl)

The feature 'sgpl' is applied to determiners which can agree with words that are either singular or plural, e.g. *dens.*

Definiteness: def (best)
indef (ubest)

Determiner Type: dem (demonstrativ)
expl (ekspletiv)
quant (kvant)
poss (possessiv)

Summing up, the features for determiners should be given in the following order: gender, number, definiteness and determiner type.

Examples:

den	Det_mf_sg_def_dem
dette	Det_n_sg_def_dem
dens	Det_mf_sgpl_def_poss

A few determiners with the feature 'quant' must also have special features in the fourth column field. The feature FREQUENT_AS_COMPOUND allows words shorter than four letters to be accepted in compounds, which is necessary for

complex numbers. The cardinal numbers between one and nine which have less than four letters receive this feature: *en, ett, to, tre, fem, sju, syv* and *ni.* All numbers between one and nine must in addition have the feature 'one' in order to allow for correct compound analysis.

Examples of fourth column field:

tre Det,pl,indef,quant,one,FREQUENT_AS_COMPOUND
fire Det,pl,indef,quant,one

Affixes

There should be two separate files for affixes, one for prefixes and one for suffixes. These files must have the regular SCARRIE intermediate dictionary format. The affixes are used by SCARRIE's compound analyzer, which analyzes productive derivations. They are allowed in derivations but not as separate words. Since some affixes require style replacement, there is a special code, M (for morpheme), which is used in the thrid column in place of N (for normal word). This code prevents the corrector from accepting affixes as words in running text. This code may then be combined with the C codes. For instance, MC12 in the third column means that the entry is an affix that must be replaced in styles 1 and 2. Affixes must contain relevant grammatical information in the fourth column. Since compounds receive their grammatical features from their last consituent, only the suffix file needs to have grammatical features in the sixth column.

Prefixes have the category Pref. There is only one feature for prefixes: the feature 'o' is assigned to all prefixes that end on the letter 'o'.

Suffixes receive the category and features that derivations ending with them have. In addition each suffix has two extra features. The first is 'suf' for suffix. The second is a feature for suffix type. The following are the suffix type features:

Suffix type:
ad
adqu
no
noad
noadqu
pre
prefi
prefo
qu
vimp
vimpad
vimpno
vinfad
vinfno

The following table shows what categories the different suffix type features may combine with and gives examples of each suffix type.

Suffix	**Combines with**	**Example**
ad	adjectives	*-artet* as in *gråartet*
adqu	adjectives and quantifiers	*-toms* as in *halvtoms*
no	nouns	*-aktig* as in *krampaktig*
noad	nouns and adjectives	*-bygding* as in *fjellbygding*
noadqu	nouns, adjectives and quantifiers	*-bent* as in *tobent*

pre	prepositions	*-settelse* as in *påsettelse*
prefi	prefixes	*-takse* as in *paratakse*
prefo	prefixes ending on -o	*-fil* as in *bibliofil*
qu	quantifiers	*-tonner* as in *tusentonner*
vimp	imperative form of verbs	*-bar* as in *studerbar*
vimpad	imperative form of verbs and adjectives	*-ing* as in *spilling*
vimpno	imperative form of verbs and nouns	*-eri* as in *spilleri*
vinfad	infinitive form of verbs and adjectives	*-voren* as in *kranglevoren*
vinfno	infinitive form of verbs and nouns	*-messig* as in *spillemessig*

For prefixes, the category Pref is listed first, followed by a comma and the feature 'o' where appropriate. For suffixes the order is as for the other categories and their features, followed by the feature 'suf' and finally the suffix type feature.

Examples of fourth column field:

astro	Pref,o
kombi	Pref
akteren	N,sg,def,suf,qu
skjegga	N,pl,def,suf,ad
teker	N,pl,indef,suf,prefo

2. Written norms in Norwegian

Norwegian has two distinct written languages, Bokmål and Nynorsk. Only Bokmål is handled in the current SCARRIE project. Within Bokmål, however, there are several written norms, ranging from "conservative" to "radical", including an official textbook norm (læreboknormalen) required in writings by government officials and in publicly approved textbooks. It is common in newspapers to find both conservative and radical variants, and they are considered to be well educated when used consistently. Written norms differ with respect to gender system (see Appendix 1) and the use of variant spellings, resulting in several combinations of these parameters. As we will show below, we will for practical purposes define only five distinct written norms in Bokmål in the context of this project.

When there are several word forms that express the same morphosyntactic word, it is possible that some of the word forms should be marked as belonging to a certain written norm. The different word forms may either be alternative spellings of the same lemma, or they may be various inflections of the same stem. (N.B. In this respect, it is inconsistent that NKL treats *aust* and *øst* as two alternative spellings of the same lemma, whereas *grease* and *gris* are treated as two unrelated lemmata.)

In the case of variant spellings of the same lexeme, there may either be alternate (*jamstilte* or *likestilte*) forms, or there may be main forms and side forms (*hovedformer* and *sideformer*). The main forms are allowed in all kinds of writing, and they are required in the textbook norm. The side forms are spellings and inflections that are not allowed in the textbook norm, but which may be used for instance in the written work of pupils in the schools. The side forms are also called bracket forms (*klammeformer*) because of the common practice of listing them in dictionaries (including Bokmålsordboken, abbreviated BMO) in square brackets.

Side forms will in general have a stylistic nuance which identifies them as belonging to a more conservative or more radical written norm than that of the main forms.

They will often be conservative, but not always. Some examples of side forms that are radical are:

raud

greidd

greidde

komma (infinitive)

Although side forms are coded consistently with brackets in BMO, their stylistic value is not coded. This must therefore be done manually.

Some forms are unapproved. Examples are:

(**II due**) se *duge*

(**duffelcoat**) se *dyffelcoat*

(**grease**) se *I gris*

(**have**) se *hage*

(**II pel**) se *pæl*

(**pelme**) se *pælme*

(**peu à peu** el. **peu om peu**) se *pø om pø*

(**sne**) se *snø*

(**syv**, **syvende**) se *sju, sjuende*

(**II torv**) se *torg*

Two of these examples are loanwords which have foreign spellings alongside of Norwegian ones. The others are conservative forms. It is doubtful if any unapproved forms are radical. We haven't found any by browsing in BMO.

Although some gender systems (see section 3 below) are closely tied to certain written norms, this is not necessarily the case. Radical Bokmål must have a three gender system, and conservative Bokmål must have a two gender system. Most writers of Norwegian, however, write neither radical nor conservative Bokmål. They either write within the textbook norm, or they write something that we might call 'neutral' Bokmål, for lack of a better term. A neutral style tends to be associated with a 2.5 gender system (see below). The textbook norm does not dictate any particular gender system, although some nouns must obligatorily be declined in the feminine, which means a pure 2-gender system is not possible there.

Side forms must always be coded in such a way that they may be replaced if they are used in a written norm to which they do not belong. For instance, in conservative Bokmål,

*Han kjøpte en **raud** rose til moren sin.*

should be corrected to

*Han kjøpte en **rød** rose til moren sin.*

In order to do this, *raud* must have a written norm feature that instructs the system to replace it under the given written norm. Replacing conditionally under a given norm (called *'style'* in the CORRie documentation) can be done by using a C (followed by the number(s) of the style(s) the replacement is valid for) in the third field and a replacement in the fifth field:

raud	...	C2	...	rød	...

Similarly, for anything but conservative Bokmål:

sne	...	C134	...	snø	...
sneen	...	C134	...	snøen	...

To sum up, there are four basic statuses mentioned in Bokmålsordboka. A word is either:

an equivalent form (*likestilt form*),

a main form (*hovedform*),

a side form (*sideform*), or

an unapproved form (*ikke-tillatt form*).

There are, however, only three different ways of coding these in BMO. Side forms are listed in square brackets, and unapproved forms are in parentheses. But equivalent forms and main forms are not distinguished from each other in any way. In addition, sometimes equivalent forms of what are obviously the same word, like *annerledes* and *annleis*, are not related to each other in any way (other than *annleis* having *annerledes* as its definition). It should, however, be possible to identify equivalent forms by obtaining lists of them, for instance from Språkrådet.

Unapproved forms appear to be of two kinds: conservative forms, basically Riksmål (e.g. *sne, syv,* cross-referenced with approved forms), or spellings unapproved for other reasons, e.g. English-based spellings (cross-referenced to the correct spelling). The latter type of unapproved forms should be corrected to the approved spellings, whereas the conservative forms should only be allowed in conservative Bokmål.

Side forms always have a radical or conservative value. Some of them may therefore only be acceptable in radical or conservative Bokmål, but many are also acceptable in a neutral style.

The fact that two forms formally have equivalent status does not mean that they are equally acceptable in all written norms. Although many equivalent forms have a neutral value, many of them also have radical or conservative values. For instance, although *mage* and *mave* are equivalent forms, *mage* may be used in any written norm, while *mave* is possible only in conservative Bokmål.

Mnemonic codes for classification of word forms

The following is a classification of forms in terms of mnemonic codes, but these codes are **not** part of the SCARRIE word form list. Instead, the word form list uses numeric codes indicating under which style certain word forms need to be accepted or replaced (see below).

RR a form so radical that it is only appropriate in radical Bokmål *(tru, greidde)*

RN a form that may be used in neutral Bokmål but that has a radical nuance that makes it inappropriate in conservative Bokmål *(steik)*

CC a form so conservative that it is only appropriate in conservative Bokmål *(mave, kuen)*

CN a form that may be used in neutral Bokmål but that has a conservative nuance that makes it inappropriate in radical Bokmål *(røke)*

N a form that is so neutral that it may be used in any style *(lete)*

U a form that is not appropriate in any written norm, for instance because of a foreign spelling *(YAP)*

T a form that is not appropriate in any written norm except the textbook norm *(fordyping, osken)*

H a hybrid form that is acceptable in neither radical nor conservative Bokmål *(melka)*, but may be acceptable in neutral Bokmål and/or the textbook norm (will be further investigated)

N.B. If there is doubt about the appropriate coding for a word form, it should be coded as more neutral. For instance, if the choice is whether a form is RR or RN, the latter should be chosen.

Written norms for SCARRIE users and consequences for allowing or replacing word forms

In SCARRIE we will explicitly allow for the following written norms, coded by their respective numbers in the SCARRIE word form list.

1. Neutral: allows side forms RN, CN and N, equivalent forms RN, CN and N, main forms RN, CN and N, no unapproved forms, 2.5 genders (allow either *boken* or *boka*).

2. Conservative: allows side forms CC, CN and N, equivalent forms CC, CN and N, main forms CC, CN and N, unapproved forms CC, 2 genders (replace *boka* with *boken*).

3. Radical: allows side forms RR, RN and N, equivalent forms RR, RN and N, main forms RR, RN and N, unapproved forms RR, 3 genders (replace *boken* with *boka*).

4: Textbook norm: no side forms, all equivalent forms, all main forms, T forms, no unapproved forms, any combination of gender systems (allow either *boken* or *boka*).

5: No written norm: allows all forms except U *(YAP)*, any combination of gender systems.

Behavior of the various dictionary forms under chosen written norms

<u>equivalent forms</u>

aske	N	should not be replaced in any style
mjølk	RN	should be replaced in 2 by *melk*
tru	RR	should be replaced in 1 and 2 by *tro*
frem	CN	should be replaced in 3 by *fram*
mave	CC	should be replaced in 1 and 3 by *mage*

<u>main forms</u>

bar (Adj)	N	should not be replaced in any style

røyke	RN	should be replaced in 2 by *røke*
deltaing	RR	should be replaced in 1 by *deltakelse* and in 2 by *deltagelse*
slukke	CN	should be replaced in 3 by *slokke*
??	CC	should be replaced in 1 and 3

side forms

mørr	N	should be replaced in 4
??	RN	should be replaced in 2 and 4
lærerer	RR	should be replaced in 1, 2 and 4 by *lærere*
deltager	CN	should be replaced in 3 and 4 by *deltaker*
??	CC	should be replaced in 1, 3 and 4

unapproved forms

sne	CC	should be replaced in 1, 3 and 4 by *snø*
atstadig	RR	should be replaced in 1, 2 and 4
YAP	U	should be replaced in all styles by *japp*

No examples of main or side forms with the status CC, of side forms with the status RN or CC, or of unapproved forms with the status RR have been found. They are nonetheless included in the list (with double question marks) since they are logically possible, and since it is clear how they should be coded if they are found.

We refer to Appendix 1 for examples of word form entries showing the conditional replacement under given written norms.

3. Gender systems for Norwegian

Norwegian has three genders, masculine, feminine and neuter. Gender is marked by combination with different determiners and adjective forms, and by having different inflections for the definite forms. However, there are alternative written standards which do not exploit all genders fully; they may not use the feminine gender and thereby use only two of the three genders. There is no official norm enforcing consistency of the gender system. As long as there is gender agreement within each individual noun phrase, an author may choose between masculine and feminine gender at random throughout a text. However, most authors prefer some consistent system among the possibilities outlined below.

In radical Bokmål, a three gender system is preferred. In conservative Bokmål and some dialects one may use a two gender system where feminine and masculine are collapsed, so that feminine nouns have the same inflection as the masculine nouns (-en) and are combined with the masculine determiners and adjective forms from the three gender system. In addition, there is a possibility to use a 2.5 gender system where feminine has its own inflection and postposed determiners but has masculine preposed determiners and adjective forms. The reverse combination (masculine

inflection but feminine preposed determiners and adjectives) does not occur and will not be treated. The situation of the feminine in the three different systems is sketched in the following table for the feminine noun *bok*:

3 gender system	2.5 gender system	2 gender system
*en (liten) bok	en (liten) bok	en (liten) bok
ei (lita) bok	*ei (lita) bok	*ei (lita) bok
boka (mi)	boka (mi)	*boka (mi)
*boken (min)	boken (min)	boken (min)

The combination of the feminine determiner with the masculine adjective form or vice versa is ungrammatical in all three systems:

**ei liten bok, *en lita bok*

Likewise, the combination of the feminine definite noun with the postposed masculine possessive determiner or vice versa is ungrammatical in all three systems:

**boka min, *boken mi*

These gender systems show some complications. Actually only the three gender system is completely consistent: those writers who use the feminine indefinite article *ei* always use the feminine inflection of feminine nouns. In the two gender system, most writers use masculine inflections (-en) for feminine nouns, but they may also use feminine inflections of a few special feminine nouns. According to Helge Sandøy, one might want to allow all the nouns that were obligatorily feminine before 1981; this list includes several hundred words. According to Helge Dyvik, however, the number of nouns actually used this way in a two gender system is extremely limited *(geita, kua, øya, jenta, hytta).* These speakers do not use the feminine possessive (e.g. *mi*), so although they may write *hytta,* they avoid writing *hytta mi,* choosing instead *hytten min* or a preposed possessive *min hytte*.

The most complicated system is the one we have called the 2.5 gender system. This system does not use the feminine indefinite article, but it does use feminine inflections of feminine nouns. It is not the case, however, that masculine inflections of feminine nouns may simply be replaced automatically by feminine inflections, i.e. *boken* is replaced by *boka.* This is because many (according to Helge Dyvik, most) writers of Bokmål use both feminine and masculine inflections of feminine nouns. In the various possible mixtures, some internal preference factors play a role. Consider e.g.:

?Mora mi skal lese boken min.

Moren min skal lese boka mi.

The first of these examples is odd, but the second is perfectly all right in this written norm. It seems that each individual noun is more or less likely to occur with the feminine suffix, and there is an implicational relation between these uses. Since *boka* is more common than *mora,* the use of *mora* implies the use of *boka,* but not vice versa.

In a chapter written by Jon Erik Hagen in his forthcoming book *PRISME-grammatikken: Norsk grammatikk for norsk som andrespråkslærere,* we found feminine nouns inflected in the following mixture of masculine and the feminine inflection:

masculine inflection

avisen
betydningen
bøyningsformen
formen
fortellingen
forutsetningen
framtidigheten
generalprøven
handlingen
imperativformen
jorden
kjernen
klassen
klokken
konteksten
motsetningen
oppstillingen
oversetningen
presensformen
preteritumsformen
setningen
teksten
verbalhandlingen
verbklassen
ytringen

feminine inflection

avisa
boka
døra
fortida
framtida
gruppa
jenta
leksa
nåtida
pila
ringeklokka
samtida
rypa
undergruppa

(Of these, only *jente* and *rype* are on the list of obligatory feminines in the textbook norm according to St. Meld. nr. 100, 1981.)

The variation among the gender systems described above is handled for the most part by having only one lexicon but three different grammars for the three different gender systems. This means that feminine nouns are always defined as 'f' in the lexicon, but the grammar will treat 'f' as 'm' in the appropriate syntactic positions when there is not a full three gender system. In addition, for definite nouns, there are different lexical entries (*boka* vs. *boken*). These can be conditionally replaced under the given norm.

References

Faarlund, Jan Terje, Svein Lie and Kjell Ivar Vannebo, 1997, *Norsk referansegrammatikk,* Universitetsforlaget, Oslo.

Hagen, Jon Erik, to appear 1999, *PRISME-grammatikken: Norsk grammatikk for norsk som andrespråkslærere,* Gyldendal, Oslo.

Landrø, Marit Ingebjørg and Boye Wangensteen, 1993, *Bokmålsordboka*. (second edition), Universitetsforlaget, Oslo.