

Dansk Sprognævn

Linguistic Diversity and Language Change - Future Challenges for MT

Sabine Kirchmeier-Andersen
Danish Language Council

sabine@dsn.dk

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Outline

- Background info
- Policy trends and language planning - Europe and the Nordic countries
- Coverage of languages with regard to LT and MT
- Trends in language change
- Conclusions

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Background info

Danish Language Council (Danish Ministry of Culture)

- monitoring the development of the Danish language
- providing the orthographic standard for schools and public institutions
- advice on language use in private and public institutions

Nordic Language Coordination (Nordic Council of Ministers)

- coordinating research, teaching and information initiatives
- strengthening the mutual understanding in the Nordic countries via the three Scandinavian languages Danish, Norwegian and Swedish

European Federation of National Institutions of Language (EFNIL) www.efnil.org

- promoting linguistic diversity
- monitoring language development and language policy

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (ef. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Main tasks of the Danish Language Council

Providing advice and information about Danish

- answering questions from the public and the authorities about the structure and usage of Danish
- publishing writings about Danish (including the council's home page and other sites), www.dsn.dk, www.sproget.dk
- providing information on language issues in newspapers, radio and tv
- cooperating with terminologists, lexicographers, and others

Monitoring the development of Danish

- collecting examples of new words and word usages
- tracking changes in orthography, morphology, syntax and semantics
- research projects on new developments in the language
- contrastive studies with other languages especially Scandinavian languages

Producing the official dictionary of Danish orthography

Dansk Sprognævn

Policy trends and language planning
in Europe and the Nordic countries

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Types of European Languages

- Indigenous minority languages
- Polycentric minority languages
- National languages
- Contact languages
- Global languages

International Language Policies

Language and language policies are discussed not only in Denmark but throughout the world. Language declarations have been formulated in UNESCO, the EU, and the Nordic Council of Ministers.

Common elements of these declarations are:

- A growing concern for the national languages as active bearers of the national culture in all domains
- The recognition of English as a working language and practical tool of communication in certain situations
- The protection of minority languages (and their speakers)
- The vision that all citizens should be enabled to learn the foreign languages that they need the most
- The vision that the education systems should support multilingual communication and intercultural competence (mother tongue + 2)



retskrivning sb., -en, -er, f
 sms, retskrivnings-, fx ret-
 skrivningssystem.
 rig (el. retlig) adj., -t.
 læge sb., -n, -r.
 lægeråd sb., -et, rets-
 egeråd, bf. pl. -ene.
 slærd adj., itk. d.s.
 slos adj., -t.

retskrivning sb., -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) adj., -t.
retslæge sb., -n, -r.
retslægeråd sb., -et, rets-
lægeråd, bf. pl. -ene.
retslærd adj., itk. d.s.
retsløs adj., -t.

The status of the Nordic languages

State constituting languages

- Danish, Finish, Icelandic, Norwegian (Bokmål and Nynorsk), Swedish

Society constituting languages

- Danish, Finish, Faroese, Greenlandic, Icelandic, Norwegian, Sámi, Swedish

Minority languages

- Meänkieli, Kven, Romani, Yiddish, German, sign languages

Nordic declaration on language policy (2006)

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Nordic Language Convention

Adopted June 17th 1981 ensures the right of Nordic citizens to use their own language in any other Nordic country (ratified by Denmark, Norway, Sweden, Finland and Iceland).

Concerns oral and written contact with state authorities such as courts, social, health, labor, tax, police and school authorities.

Covers translation and interpretation and permanent special services (in areas where there is a strong representation of speakers of a another Nordic language).

Expenses paid by the states.

Multilingualism and Plurilingualism

- The Nordic region considers itself as plurilingual and multilingual.
- Sámi, Kven, Meänkieli, Romani, Yiddish, Finnish (in Sweden) and German (in Denmark) are all official minority languages.
- There is a special obligation to protect and support Greenlandic.
- More than 200 other languages are spoken in the Nordic region. It is highly desired to establish expert groups that can give advise on these languages or refer to similar expert groups in Europe.

Nordic declaration on language policy (2006)

retskrivning sb., -en, -er, f
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (ef. retlig) adj., -t.
retslæge sb., -n, -r.
retslægeråd sb., -et, rets-
lægeråd, bf. pl. -ene.
retslærd adj., itk. d.s.
retsløs adj., -t.

Parallism between English and the Nordic languages

- the Nordic languages as well as English should be used as languages of science,
- publishing scientific results in one of the Nordic languages should be as highly regarded as publishing in English,
- courses in scientific terminology and LSP should be taught both in English and in one of the Nordic languages,
- universities and other places of higher education should develop long term strategies for the choice of languages, language parallelism, language training and translation,
- the Nordic terminology institutions must continue to coordinate terminology in the emerging domains,
- companies and industrial organisations must be encouraged to develop strategies for language parallelism.

Nordic declaration on language policy (2006)

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, *fx* ret-
 skrivningssystem.
 retslig (el. retlig) *adj.*, -t.
 retslæge *sb.*, -n, -r.
 retslægeråd *sb.*, -et, rets-
 lægeråd, *bf.* pl. -ene.
 retslærd *adj.*, *itk.* *d.s.*
 retsløs *adj.*, -t.

Language shifts

Number of speakers	Number of languages (6095)
More than 100 mio.	8
10-99 mio.	72
1 mio. - 9,9 mio.	239
100.000 - 999.999	795
10.000 - 99.000	1605
1000 - 9999	1782
100 - 999	1075
10 - 99	302
1 - 9	181

Ethnologue

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Critical factors for language survival

- Number of speakers
- Number of young speakers
- Writing tradition
- Use of the language in education
- Status of the language
- Status of the language users
- **Access to language technology**

David Crystal: Language Death 2000

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Is there a need for LT and MT?

Of course!!

Enhancing communication

Enabling more efficient communication

Enabling information retrieval

Enabling communication with other languages

Enabling true parallelism

Keeping small languages attractive on the international market

- products and product descriptions/manuals

Keeping people proud of their language

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (ef. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf.* pl. -ene.
retslærd *adj.*, itk. *d.s.*
retsløs *adj.*, -t.

But even much more basic:

- there is a general need for better support of the different languages on the web and in all it-applications.

We still find products especially on the web that cannot cope with our national characters

æ, ø, å => ae, oe, aa

They can be used in URL's to some degree (but this is not widely known)

- but they can still not be used in email-addresses.

Århus => Aarhus

Keeping people proud of their language?

Dansk Sprognævn

Coverage of languages with regard
to Language Technology and MT

	eng	fra	deu	spa	ita	por	nld	swe	ell	pol	dan	ces	fin	ron	hun	bul	slv	lav	lit	slk	est	mlt	gle
eng	1320	109	111	107	100	84	50	30	24	35	13	11	15	10	16	10	10	7	5	5	4	3	2
fra	109	847	79	66	52	41	36	19	22	14	11	9	9	8	8	7	8	7	5	5	4	3	2
deu	111	79	720	42	38	26	20	18	14	18	12	10	10	9	10	7	8	7	5	5	4	3	2
spa	105	65	40	650	35	29	19	17	14	11	12	9	9	8	8	7	8	5	5	5	4	3	2
ita	100	52	38	36	599	25	19	16	14	11	11	9	9	8	8	7	8	5	5	5	4	3	2
por	85	41	26	29	25	497	18	15	13	11	11	9	9	8	8	6	8	5	5	5	4	3	2
nld	49	36	20	19	19	18	376	16	14	10	11	9	9	8	8	6	8	5	5	5	4	3	2
swe	30	17	18	17	16	15	16	272	13	8	12	9	10	8	8	6	8	5	5	5	4	3	2
ell	23	22	14	14	14	13	14	13	267	7	9	7	8	7	6	7	6	5	5	3	3	3	2
pol	35	14	18	11	11	11	10	8	7	250	7	9	8	7	7	6	7	7	5	4	3	3	2
dan	13	11	13	12	11	10	11	12	9	7	181	7	8	7	7	6	7	5	5	4	4	3	2
ces	10	9	9	9	9	9	9	9	7	9	7	168	9	8	8	7	7	6	5	4	3	3	2
fin	16	9	10	9	9	9	9	10	8	8	8	9	157	7	7	6	7	5	5	4	3	3	2
ron	10	8	9	9	8	8	8	8	7	7	7	8	7	144	7	7	7	4	4	4	3	2	1
hun	15	8	9	8	8	8	8	8	6	7	7	8	7	7	129	5	8	5	5	5	4	3	2
bul	9	7	7	7	7	6	6	6	7	6	6	7	6	7	5	151	5	4	4	2	2	2	1
slv	10	8	8	8	8	8	8	8	6	7	7	7	7	7	8	5	112	5	5	5	4	3	2
lav	7	7	7	5	5	5	5	5	5	7	5	6	5	4	5	4	5	93	5	3	3	3	2
lit	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	4	5	5	75	3	3	3	2
slk	5	5	5	5	5	5	5	5	3	4	4	4	4	4	5	2	5	3	3	5	4	3	2
est	4	4	4	4	4	4	4	4	3	3	4	3	3	3	4	2	4	3	3	4	4	3	2
mlt	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	3	3	3	3	3	2
gle	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2

ts-

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (ef. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Resources available for English ->Danish

[ESTeam Translator](#) *eng* → *dan*

[Google Translate](#) *eng* → *dan*

[Interpret](#) *eng* → *dan*

[TranslationBooth](#) *eng* → *dan*

[InterTran Web Site Translation Server](#) *eng* → *dan*

[GramTrans](#) *eng* → *dan*

[LinguaNet](#) *eng* → *dan*

[Language Weaver Enterprise Translation Server](#) *eng* → *dan*

Parallel Corpora

[EuroParl release 3.0](#)

[OPUS; EUconst](#)

[OPUS; KDEdoc](#)

[OPUS; KDE](#)

[Acquis Communautaire](#)

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

LT status for the Nordic countries

At present the largest investments in LT for the Nordic countries are made in tools for Greenlandic and Sámi, but many basic resources and tools for the main Nordic languages are still lacking

BLARKs do not exist for all languages

A growing awareness of the need for LT infrastructure

Term banks and text banks are being developed for Norwegian and Swedish financed by the state and made publicly available

Denmark is developing resources and tools in DK-CLARIN. 10 mil. DKK (1,4 mil. Euro) (general language corpus + LSP corpus + infrastructure)

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Status

Governments still believe that language challenges should be solved on a commercial basis and left to market mechanisms...

The facts are:

- local LT industry is extremely fragile
- small languages are not considered attractive markets for international companies

Dansk Sprognævn

Trends in language change

retskrivning sb., -en, -er, i
sms. retskrivning
skrivning
retslig (el. retlig) adj., -t.
retslæge sb., -n, -r.
retslægeråd sb., -et, rets-
lægeråd, bf. pl. -ene.
retslærd adj., itk. d.s.
retsløs adj., -t.

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Language change - two definitions

Language shifts - A community of speakers of one language shifts to using another language (Fishmann, Romaine, Crystal)

Change within the language

- new words
- new constructions
- new senses

retskrivning sb., -en, -er, i sms. retskrivnings-, fx retskrivningssystem.
 retslig (ef. retlig) adj., -t.
 retslæge sb., -n, -r.
 retslægeråd sb., -et, retslægeråd, bf. pl. -ene.
 retslærd adj., itk. d.s.
 retsløs adj., -t.

Language change due to changes in the communicative setting

One to one	Letters => phone, email, sms
One to many	Books => newspapers, magazines, webpages
Many to many	=> chat, blog, Facebook, Twitter ...



retskrivning sb., -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (ef. retlig) adj., -t.
retslæge sb., -n, -r.
retslægeråd sb., -et, rets-
lægeråd, bf. pl. -ene.
retslærd adj., itk. d.s.
retsløs adj., -t.

Development of neologisms

New words

30.000 new words in Danish each month (20 mio. words, newspaper articles)

1000-2000 actual new words in general language - 12.000 new words pr. year

Compounds (language technology => sprogteknologi)

Acronyms (GPS, MT)

Loan words (Blow out, tweet, fuck)

(fuck: 1,1 % of all newspaper articles in 2009, sgu: 2,5 %)

Domesticated words (Hockey mom - hockeymor, nerd => nørd)

Translated expressions (at the end of the day - ved dagens slutning)

Mixtures

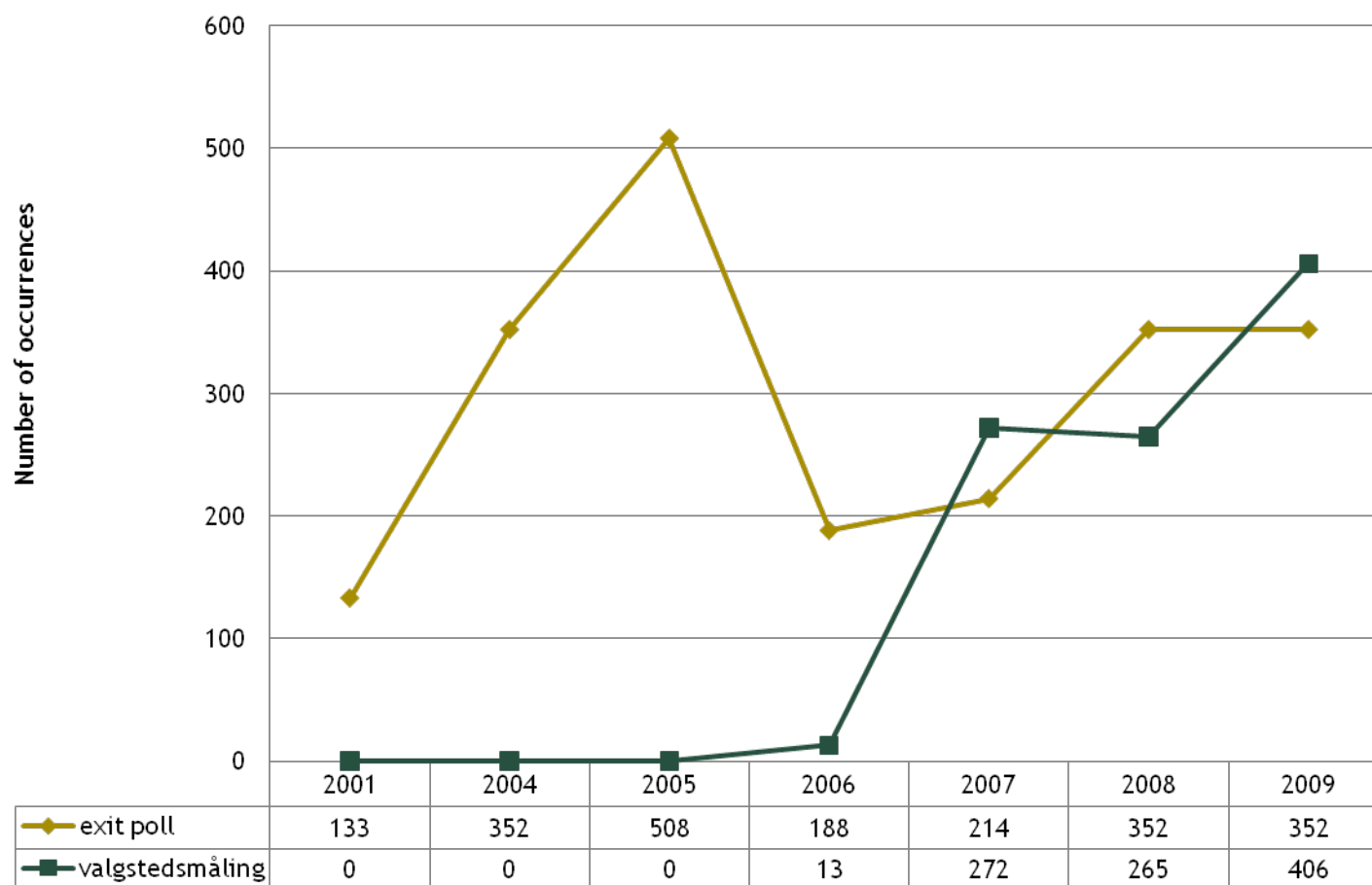
New word-patterns - new senses of existing words

stand up for X - stå op => stå op for X

hang out with X - hænge X ud => hænge ud med X

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
 skrivningssystem.
 retslig (el. retlig) *adj.*, -t.
 retslæge *sb.*, -n, -r.
 retslægeråd *sb.*, -et, rets-
 lægeråd, *bf. pl.* -ene.
 retslærd *adj.*, *itk. d.s.*
 retsløs *adj.*, -t.

Negotiation of norms in language communities



retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf.* pl. -ene.
retslærd *adj.*, *itk.* d.s.
retsløs *adj.*, -t.

Status for language change

Language change and language shift exist - and have always done.

However, we do not now

- how fast changes take place today
- to which degree the new forms of communication speed up these changes
- how changes are negotiated in the social media

Conclusions

Dansk Sprognævn

retskrivning sb., -en, -er, i
sms. retskrivning for ret-
skrivning, sprogretskrivning
retslig (el. retlig) adj., -t.
retslæge sb., -n, -r.
retslægeråd sb., -et, rets-
lægeråd, bf. pl. -ene.
retslærd adj., itk. d.s.
retsløs adj., -t.

Conclusions

Language diversity exists - but is under pressure even for national languages

- language technology and especially MT can contribute to improving the situation
 - by enabling language users to continue to use their own language
 - by enhancing the status of the language
- However, for the development of language technology the existence of language resources - first and foremost text and speech corpora is absolutely crucial.
- Especially for smaller languages these resources have to be freely available to stimulate development of products
- Copyright laws and other legal obstacles make it impossible to give free access to corpus resources

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.

Solution

EU should move towards a general revision of copyright laws that gives free access to corpus resources for LT projects under certain conditions, for instance via national libraries - and EU should put pressure on all member countries to adopt these revisions.

Conclusions

Language change and language shifts exist

- but many LT products are not able to adapt to these changes

Governments (and the EU) are still under the expression that LT projects and the compilation of language resources are singular events: You build a system and then the problem is solved - at least for some time.

Maintaining linguistic infrastructure should rather be conceived as continuous effort comparable to securing electricity supplies

retskrivning *sb.*, -en, -er, i
sms. retskrivnings-, fx ret-
skrivningssystem.
retslig (el. retlig) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf.* pl. -ene.
retslærd *adj.*, itk. *d.s.*
retsløs *adj.*, -t.

Solutions

- Future projects must demonstrate how the products are kept up to date with regard to language change.
- Research projects should investigate the nature of language changes in order to provide the necessary knowledge for both industry and governments
- In order to measure whether our efforts have an effect - we need a general methodology for monitoring the development of the status and the corpus of languages - a European language monitor.

Dansk Sprognævn

Thank you!
Vielen Dank!
Mange tak!

retskrivning *sb.*, -en, -er, *i*
sms. retskrivning
retskrivning
retslig (*el. retlig*) *adj.*, -t.
retslæge *sb.*, -n, -r.
retslægeråd *sb.*, -et, rets-
lægeråd, *bf. pl.* -ene.
retslærd *adj.*, *itk. d.s.*
retsløs *adj.*, -t.