

Collecting language resources from the private sector in Norway

Jon Arild Olsen
Senior adviser
National Library of Norway
jon.olsen@nb.no



Some background

- Norway is not a EU member: no language resources from the EU administration.
- All current resources in eTranslation come from Norwegian public agencies:
 - ✓ The Norwegian Ministry of Foreign Affairs
(ca. 700 000 units)
 - ✓ The Norwegian Consumer Council
(ca. 300 000 units)



Some more background

- Few public agencies have in-house translation services using CAT tools.
- 80% of public sector translations are outsourced.
- 5 companies dominate the market (but subcontracting is very common).



Norwegian language policy

- Two official written languages: Bokmål and Nynorsk.
- The majority (ca. 85%) uses Bokmål.
- Norwegian language policy ensures equitable treatment of both official language forms:
 - ✓ Citizens addressing themselves to public administration must receive written answers in their chosen language form.
 - ✓ Text books for primary and secondary schools must be available in both language forms.
 - ✓ Idem for forms used by public agencies.
 - ✓ Both official language forms must be represented by at least 25% of publications produced by public agencies.



Norwegian language policy and eTranslation

- The Norwegian ministry of Culture (in charge of language policy) has decided that both official forms of Norwegian must be available in eTranslation.
- Challenge: Very little public sector translation between Nynorsk and English.



Priorities for Norwegian language in eTranslation

- Increase the number of bilingual resources in English - Norwegian (Bokmål or Nynorsk).
- Collect/create a parallel corpus in Norwegian Bokmål and Nynorsk.



Contributions from the private sector

- The National Library has decided to address commercial companies producing bilingual resources.
- So far, three agreements have been made.



1. Amesto Translations Ltd.

- Norway's largest commercial translation company with many public sector clients.
- Normally agreements with public agencies do not say anything about ownership or access to TM.
- The National Library has received TM from all public sector assignments since 2011, ca. 1,5 million units.
- The National Library takes responsibility for all further use.



- The National Library has informed the concerned ministries about the agreement.
- The National Library has checked the content of the resources.
- Approximately 7 % will be deleted due to privacy/confidentiality issues.
- Of the remaining resources, 60% can be used as they are; 40% requires anonymization.



2. Standards Norway

- Norway's biggest standards organization responsible for all areas except electronics and telecommunication.
- Access to translated standards is payable.
- The National Library has received TM from standards translated from English since 2012, ca. 500 000 units.
- The TM are “scrambled” and metadata indicating original position are deleted.



3. Nynorsk Press Office

- Sells news reports to Norwegian media.
- Reports translated from Bokmål into Nynorsk.
- Uses Apertium, a Spanish open-source, rule-based MT program.
- All results are «manually» quality checked.
- The National Library will receive TMs from ca. 70 000 news reports.



Future collaboration with commercial language services

- Meeting in the Ministry of Local Government and Modernization about how to ensure regular collection of TM from commercial companies working for public agencies.
- Some preliminary conclusions:
 - ✓ Impractical to address many public agencies.
 - ✓ Better to collect TM directly from a few commercial providers.
 - ✓ The National Library needs a clear mandate/authority.

Three possible solutions .

- a) Include clauses stipulating public ownership and access in each particular contract or agreement.

The Agency for Public Management and eGovernment (Difi) can provide general advice on how to conduct public tendering processes for language services.

- b) A possible Swedish solution: organize a common public tendering process for all translation and interpretation services needed by public agencies.

Requires action by the Agency for Public Procurement.

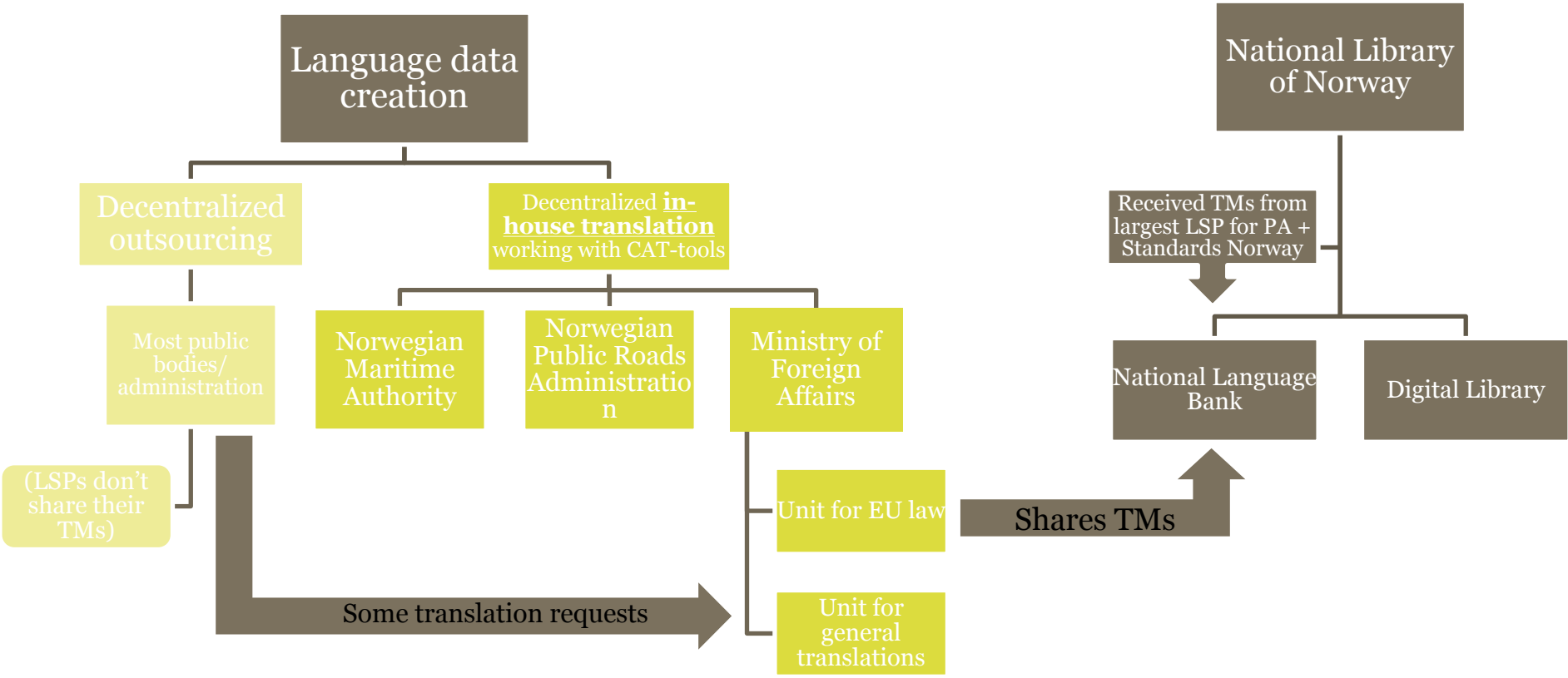
- c) Include clauses on public ownership and access to TM in laws and regulations.

Requires action on a ministerial and parliamentary level.



- The concerned parties will discuss these solutions during the forthcoming Norwegian ELRC-workshop November 20th.

Infrastructure of language data on national level in Norway_ V1.4



Thank you for your kind attention!