Speech is **Golden**

- on ASR at the service of the Danish public sector

Peter Juel Henrichsen

*DanCAST - Danish Center for Applied Speech Technology*
*Copenhagen Business School*
*University of Copenhagen*

pjh.ibc@cbs.dk
<table>
<thead>
<tr>
<th></th>
<th>Slovenia</th>
<th>Denmark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L1 speakers</strong></td>
<td>2.2 mio</td>
<td>5.5 mio</td>
</tr>
<tr>
<td><strong>EU working language</strong></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Mix of municipalities</strong></td>
<td>5-6 city / many small</td>
<td>4 city / 94 small</td>
</tr>
<tr>
<td></td>
<td>Slovene</td>
<td>Danish</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td><strong>L1 speakers</strong></td>
<td>2.2 mio</td>
<td>5.5 mio</td>
</tr>
<tr>
<td><strong>EU working language</strong></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Mix of municipalities</strong></td>
<td>5-6 city / many small</td>
<td>4 city / 94 small</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Slovene</th>
<th>Danish</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inflected language</strong></td>
<td>&gt;&gt;English</td>
<td>&gt;English</td>
</tr>
<tr>
<td><strong>Compounding</strong></td>
<td>&gt;English</td>
<td>&gt;English</td>
</tr>
<tr>
<td><strong>Rich in vowel qualities</strong></td>
<td>&gt;English</td>
<td>&gt;&gt;English</td>
</tr>
</tbody>
</table>
This talk

1. Why ASR in the municipalities?
2. ASR - the technology
3. Trough of disillusionment
4. The new alliance
Why ASR in the Danish municipalities?

The local authorities smelled a business case...

The slick salesman:

1) economy! - the medical case
2) speech aid for the challenged
3) 'welfare tech' - assisting in difficult working situations

The technology looked extremely user friendly and mature

[ad]

On top of that, the *real* world had embraced ASR already
Gartner hype curve
tazti speech recognition software
Now on Sale
BUY NOW
All NEW
Dragon Dictate for Mac, v4
However, it did not go so smoothly!

A quick intro to ASR
ASR - the technology

Three central components

- Acoustic model (AM)
- Language model (LM)
- Search engine referring to AM og LM
Acoustic model

A set of recognizers, one for each language sound ('phone')
Language model

- Unigrams
- Bigrams
- Trigrams
- (n-grams)

all frequency annotated (NB! corpus-driven)

UNIGRAMS:
the > is > in > cat > oven

BIGRAMS:
the cat > the fat >>> *the that >>> **the sat
Language model

- Unigrams
- Bigrams
- Trigrams
- (n-grams)

all frequency annotated (NB! corpus-driven)

UNIGRAMS:
the > is > in > cat > oven

BIGRAMS:
the cat  About 89.400.000 results   (by Google)
the fat  About 40.100.000 results
the that About 10.300.000 results
the sat  About 7.020.000 results

TRIGRAMS:
is in the  >> in the oven  >> ...

N-GRAMS (domain specific mwe.s)
the cat is on the mat
the cat is in the oven
ASR: a search engine over AM and LM
How are the AM and the LM developed?

Based on annotated corpus data

(tokenized, transcribed, tagged, time-coded)

Example: the speech corpus

[rec]
the cat is in the oven

<table>
<thead>
<tr>
<th>the</th>
<th>cat</th>
<th>is</th>
<th>in</th>
<th>the</th>
<th>oven</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH</td>
<td>AH0</td>
<td>K</td>
<td>AE1</td>
<td>T</td>
<td>IH1</td>
</tr>
<tr>
<td>0</td>
<td>140</td>
<td>170</td>
<td>210</td>
<td>280</td>
<td>350</td>
</tr>
<tr>
<td>680</td>
<td>810</td>
<td>1010</td>
<td>1260</td>
<td>1310</td>
<td>???</td>
</tr>
</tbody>
</table>

(phonetic script: cmu dict)

[play]
### Training data

<table>
<thead>
<tr>
<th>Acoustic model materials</th>
<th>size (order of mag.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonetic dictionary</td>
<td>100,000 lemmas</td>
</tr>
<tr>
<td>speech recordings (multi-speaker)</td>
<td>100 hours</td>
</tr>
<tr>
<td>speech recordings (focus users)</td>
<td>1 hour each</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language model materials</th>
<th>size (order of mag.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>text corpus (general)</td>
<td>100M words</td>
</tr>
<tr>
<td>text corpus (specific for professional area)</td>
<td>100k words</td>
</tr>
<tr>
<td>non-linguistic tokens (forms, symbols, ...)</td>
<td>100 documents</td>
</tr>
</tbody>
</table>

*back on track*...
Status as of 2011
ASR contracts by 2011

Vendors: 4

- KMD
- IBM Denmark
- PDC Dictus
- Max Manus

Technological suppliers: 1

Investors (municipalities): 22

Positive business-cases:
ASR contracts by 2011

Vendors: 4

- KMD
- IBM Denmark
- PDC Dictus
- Max Manus

Technological suppliers: 1

Investors (municipalities): 22
Positive business-cases: 0
Did Gartner lie?

Maybe not,

but several things conspired against the municipalities:

- Danish has difficult words: long, inflected, ...
- Danish has difficult vowels: lenitions, reductions, assimilations, ...
- MONOPOLY
The case of Odense

(they did everything right)
Odense features

- Three year budget
- Dedicated project leader
- Training programme for employees
- 900+ participants = North-Europe's biggest
Odense did everything right - and failed

- Inflated expectations - 'savings' already entered in next-year budget!
- No clear HR policy (few employees liked ASR, most gave up)
- Vendors soon vanished: Poor service after contract was signed
- Extremely slow updates (waiting 24 months for new context files)

At project end, ~150 active users (<20%)
The new alliance
The new alliance

Steering committee

- OS2 (50+ Danish municipalities, 100% flat organization)
- DanCAST (Copenhagen Business School)

Advisory board

- KOMBIT (independent advisor)
- CST (Copenhagen University)
- Danish Parliament (Folketinget)
- Danish National Broadcast (Danmarks Radio)
**First action point:** Recycling of resources

Opaque module structure
Similar experiences everywhere

- all keep paying for the same resources (e.g. phonetic lexicon)
- loss of ownership to own data (e.g. annotated speech files)
- licence lock-in (change product = begin from scratch)
- no knowledge transfer (data exchange barred)
### Recycling corpus materials

<table>
<thead>
<tr>
<th>Acoustic model materials</th>
<th>size (order of mag.)</th>
<th>recycled?</th>
</tr>
</thead>
<tbody>
<tr>
<td>phonetic dictionary</td>
<td>100,000 lemmas</td>
<td>yes</td>
</tr>
<tr>
<td>speech recordings (multi-speaker)</td>
<td>100 hours</td>
<td>yes</td>
</tr>
<tr>
<td>speech recordings (specific users)</td>
<td>1 hour each</td>
<td>no</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language model materials</th>
<th>size (order of mag.)</th>
<th>recycled?</th>
</tr>
</thead>
<tbody>
<tr>
<td>text corpus (general)</td>
<td>100,000,000 words</td>
<td>yes</td>
</tr>
<tr>
<td>text corpus (specific)</td>
<td>100,000 words</td>
<td>no</td>
</tr>
<tr>
<td>non-linguistic tokens (forms, symbols, ...)</td>
<td>100 documents</td>
<td>partly</td>
</tr>
</tbody>
</table>
Optimal recycling of training data

Recyclable : Project specific  =  50 : 1
Drawing the line between \textit{mine} and \textit{yours}
### The toolbox

(all is open domain)

<table>
<thead>
<tr>
<th>Tools (assorted)</th>
<th>Status 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCR scanner</td>
<td>OK</td>
</tr>
<tr>
<td>De-formating</td>
<td>OK</td>
</tr>
<tr>
<td>Tokenizer</td>
<td>OK</td>
</tr>
<tr>
<td>Anonymizer</td>
<td>OK</td>
</tr>
<tr>
<td>Symbols (num, abbrev, ...)</td>
<td>OK</td>
</tr>
</tbody>
</table>

(*) = needs localization
**Example:** Raw document to structured text

<table>
<thead>
<tr>
<th>VVS-arbejde</th>
<th>Godkendt</th>
<th>Arbejdsområde</th>
</tr>
</thead>
<tbody>
<tr>
<td>på grundafskar</td>
<td>Godkendelse</td>
<td>Arendrup et al.</td>
</tr>
<tr>
<td>på grundafskar</td>
<td>godkendelse</td>
<td>Arendrup et al.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Godkendelse</th>
<th>Arendrup et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Underskrivelse:**

**Dato og underskrift:**

<table>
<thead>
<tr>
<th>Navn</th>
<th>Dato</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/S Auning</td>
<td>20.02.2023</td>
</tr>
<tr>
<td>Birkværksfælles</td>
<td>00430</td>
</tr>
<tr>
<td>A/S Auning</td>
<td>00450</td>
</tr>
<tr>
<td>Navn</td>
<td>A/S Auning</td>
</tr>
</tbody>
</table>
Syddjurs Kommune
Team Byggeri
Hovedgaden 77
8410 Rande
Afdelingens hovednr: 8753 5510

Undertegnede autoriserede mester erklærer at have udført følgende arbejder efter gældende bestemmelser:

WS-arbejde
VA Vand- ogksanitetsarbejder I Gasinstallationær pé ejendommen

Ejer/bygherre
Byggelilladelsens dato Arbejdet færdigt den 09.01.2014
Undertegnede autoriserede mester erklærer at have udfart følgende arbejder efter gældende bestemmelser:

WS-arbejde
VA Vand- og sanitetsarbejder I Gasinstallationér på ejendommen

Ejer/bygherre
Byggeviladselsens dato Arbejdet færdigt den 09.01.2014
Prognose og behandlingsmuligheder:
Der er klar **overensstemmelse** mellem de objektive fund og borgers fortælling.
- Foreligger der lægeskøn eller udtalelse fra egen læge ang. **pronose**, **arb evne mm**?
Ja. Egen læge ser ikke noget arbejdsmarkedsperspektiv for **Marianne**.
Second action point: Preparing generic specs and documents

Documents available in generic/embryonic forms:

- Databehandleraftaler ("data processing agreement")
- Systemkrav ("system requirements", specs)
- Udbudsmaterialer ("bidding materials")

...
**Third action point:** Managing the bidding situation

Bidding material should specify

- clear division of components (dictionary, AM, LM)
- transparent recycling (reused vs. new data)
- only latest-state of databases can be owned by provider
- short update-cycle for AM and LM

Either as requirements or as desiderata

Bidding material should refer to the shared corpora, e.g. qualification round

These features are as important for return-of-investment as are WER etc.
The HUB

www.dancast.dk
The DanCAST hub for ASR localization support

---

**REPOSITORIUM**

<table>
<thead>
<tr>
<th>Say</th>
<th>Seby</th>
<th>Kappel</th>
<th>Oboester</th>
<th>Familiendrømme</th>
</tr>
</thead>
</table>

---

**BIBLIOTEBK**

- Værktøj
  - tokenisering
  - distilling
  - omformatering
  - ...

- Vejledninger
  - "Testimonials"
  - Forum

**NETVÆRK**

- DanCAST's forskere
- Ph.D. (erhverv/fri)
- Studenter (opgave, lønberjød?)
- Øvrige non-kommercielle interessenter
Status quo 2016

May 2015: First 100% Danish-Danish contract
Januar 2016: First independent SME established
July 2016: Now 3 Danish SME start-ups

Neither would have existed without the shared HUB-data

Bidding rounds are now more fair (results not given in advance)

Economically most significant result:

   Licence conditions are vastly improved,  
   even in old contracts!
So, a happy end?

Not entirely:
  Government cuts in 2015 and 2016 in all public budgets
  Also universities are currently firing researchers

However:
  The HUB survives and is now almost self-sustaining
By way of conclusion

If we could start over...

Step 1. Prepare the ground
   IT responsibles: Nuts-and-bolts courses
   Decision makers: Adjusted expectations!
   End-users: interest groups for employees

Step 2. Collect existing materials
   Corpora
   Tools for annotation, alignment, tagging, anonymization, ...

Step 3. Establish a data portal
   Restricted entry - for individual municipalities
   Semi-restricted entry - data-sharing among municipalities
   Unrestricted entry - fully processed and anonymized data

Step 4. Create a library of generic formulas (semi-restricted access)
   Agreements, tender material, specs

Step 5. Prepare for bidding rounds
   Organize groups of municipalities
   Require!
THE END