

Open access to sign language corpora

Onno Crasborn

Department of Linguistics / Centre for Language Studies (CLS)
Radboud University Nijmegen

NWO grant 380-70-008

Overview

- Open access to the *Corpus NGT* *Why do we want it?*
- *Creative Commons* licenses
- Ethical questions *Should we do it?*
- An outlook on the future *Will we regret it?*

Choice for open access

Large potential use beyond the scientific world

- Deaf education
- L2 teaching

Few sign teaching materials

NGT as a first language (± 5,000 deaf people)

- There is one bilingual language teaching method (2005)



NGT as a second language (± 15,000 hearing and deaf people)

- Few teaching materials (2003: first official NGT teaching method (level A0 → A2))



Other signed materials



Learning NGT

The different genres in the Corpus NGT offer possibilities for an enormous increase in NGT input for all NGT learners:

- bed time stories for small children
- the narrations and discussions by deaf adults may serve as role model input for older deaf children
- all genres can be used for various L2 learning purposes by hearing adults, but especially for training of receptive skills

Open Access to the Corpus NGT

- All movies and annotations are freely accessible through the web
- Restricted metadata are publicly available
- Scientific access: all data can be accessed by a web browser, and can be searched by their IMDI metadata description (standardised) and the available ELAN annotations (not standardised)
- General public: examples will be available via web sites
- No registration or password needed

Scientific use

General public

IMDI Browser

Web sites to be developed

Creative Commons licenses

- Goal: explicitly allow use of copyrighted material
- Three types of restrictions:
 1. Mention author's name
 2. Do not use for commercial purposes
 3. Do not modify
 - or: if you do, share under the same conditions

From 'all rights reserved' to 'please use, but don't do x and y'

CreativeCommons.org

- Authored material in a new age
- Internet allows easy sharing and easy access, but also unwanted copying and abuse
- 1998: extension of duration of US copyright laws
 - ⇒ further protection
- CC: copyright is OK, but use and reuse should be promoted rather than discouraged
- Harvard & Stanford legal experts designed a internet-adapted reference to copyright laws (2001)
- More than 20 different translations available, all taking into account national laws but staying close to the original
- Dutch version: 2004

CreativeCommons.org

- Advantages
 - the licenses refer to or extend existing national copyright laws
 - the various national translations/adaptations stay close to the US original, so that a world-wide standard is coming about
 - the promotion of such a common standard will ease the acceptance and awareness among Internet users
- Different versions
 - detailed legal text
 - detailed legal text in machine-readable form
 - layman version (the images or short texts)

CC for sign language corpora



Should derivative works be allowed (1) or not (2, ND), or only under the condition that it is published under the same conditions (3, SA)?

- Use of the work in a collection is always permitted and does not count as a derivative work, translation does. But how about using a selection of a video clip?
- You cannot prevent abuse, and it will be hard to take legal action when necessary
- General moral rights to the good name of the creator and signer still apply
- What counts as 'commercial use'?
- Specific commercial uses can still be negotiated or allowed (e.g. inclusion of clips on commercial dictionary DVDs)

License for the Corpus NGT



- BY, NC, SA for all movies and annotation files
- BY for all other documents (manuals etc.)
- A brief segment at the beginning and end of each movie clip is used to refer to the text of the license and to the project of which the movie forms a part
 - no direct reference to the authors (Crasborn, Zwitserlood & Ros)
 - no reference to the names of the signers
- For the average user reference to a special project web site (www.corpusngt.nl) would be convenient, but maintenance is hard to guarantee.

Expected future developments

- Extension of materials by others
 - additional annotations (glosses, translations, and comments) will reveal more about the content of the interaction, and make the content searchable
- Technological developments
 - improved searching and data mining tools
 - automatic sign recognition and annotation
 - face recognition software, linking content to signer's identity

Three questions on ethics

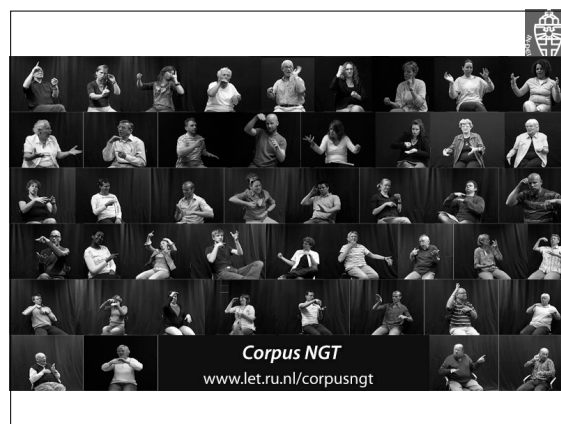
If we publish sign language movies on internet,

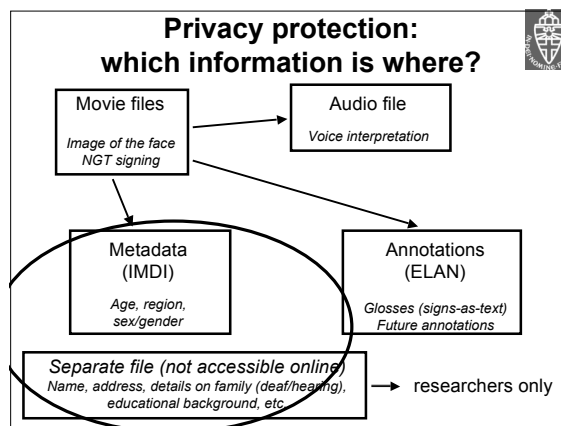
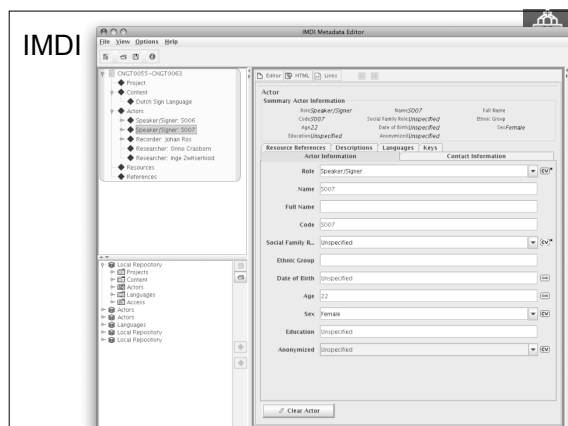
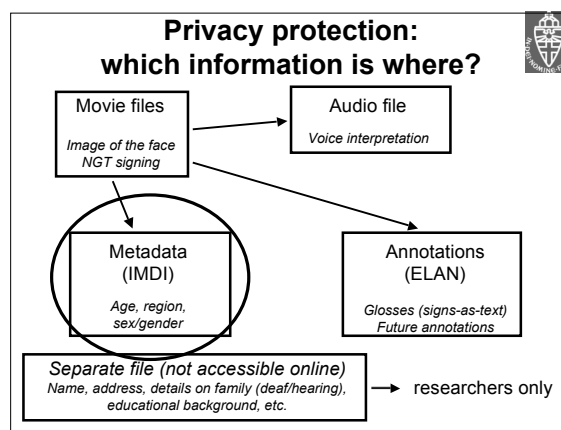
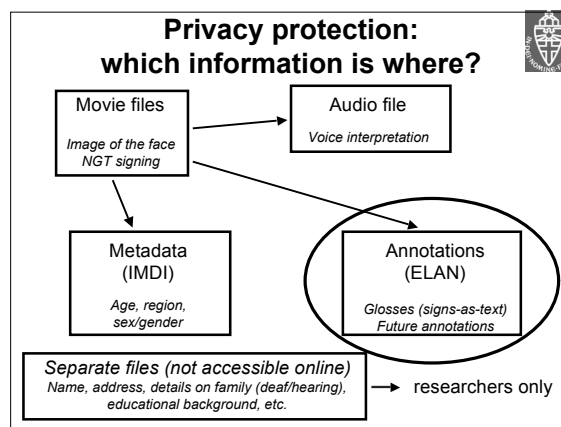
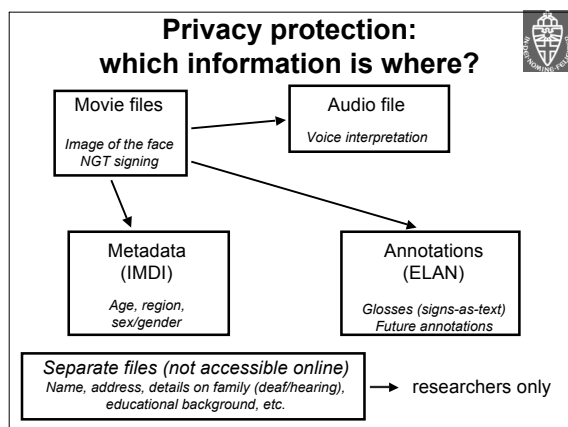
1. how can we protect the privacy of the deaf signers?
2. do people really understand (and can they foresee) the consequences?
3. can we prevent a normative effect of the publication?

Question 1. How to protect the privacy of the signers?

- *"Don't talk too much about private issues!"*
- We use a consent form
 - Do you agree to be videotaped?
 - Do you agree with the video clips being published online?
 - etc.
- Signers can withdraw consent for specific clips after seeing the DVD with their own material
- We publish as little personal information as possible
 - Included: age, region, gender/sex
 - Not included: name, acronym, address, family deafness details

But: how to protect the identity in movies?!





Protected database

Never publically accessible!

Are we doing enough to protect privacy?

- We want open access
 - The movies are unique in quality and quantity, and should be usable not only for research but for any purpose.
- ⇒ We limit the searchable (text) information as much as possible
 - (but have to offer some metadata for researchers)
- ⇒ We ask for the signer's explicit consent

What do you think?

Three questions on ethics

If we publish sign language movies on internet,

1. how can we protect the privacy of the deaf signers?
2. do people really understand (and can they foresee) the consequences?
3. can we prevent a normative effect of the publication?

Question 2: is informed consent enough?

- To give informed consent, a person must...
 - have the legal capacity (> not children)
 - give free consent (> not obliged to do so)
 - be able to make an informed decision

Question 2: is informed consent enough?

- To give informed consent, a person must...
 - have the legal capacity (> not children)
 - give free consent (> not obliged to do so)
 - be able to make an informed decision
- Do people understand what publication on internet means?
 - Always available to the whole world
 - Impossible to withdraw (due to the nature of internet, and because of our limited manpower)
- Can anyone foresee the consequences?
 - I cannot!

Publishing images of people



⇒ Someone like Onno will not happy in 10 years time.

Internet & technology

- Can anyone think of what the Internet will be like in 10 years time?
Did you ever dream of Google Earth 10 years ago?!

Internet & technology

- Can anyone think of what the Internet will be like in 10 years time?
Did you ever dream of Google Earth 10 years ago?!
- What will happen to video technology?
Face recognition built in to Windows?

Three questions on ethics

If we publish sign language movies on internet,

1. how can we protect the privacy of the deaf signers?
2. do people really understand (and can they foresee) the consequences?
3. can we prevent a normative effect of the publication?

Question 3: can we prevent a normative effect?

- Recording these and not other people. How special are these 100 people?
"Why was I not asked to participate?"
"These hearing university researchers always ask the same people!"
- Is this "NGT"?
To what extent do elements of Dutch belong in the language?
Older signers show more influence of Dutch than younger signers.
- Risk of a normative effect
Interpreters may see comprehension of these people as their ultimate diploma.
Teachers and parents may compare children to this group.

Corpus NGT: three questions on ethics

If we publish sign language movies on internet,

1. how can we protect the privacy of the deaf signers?
2. do the people we recorded really understand (and can they foresee) the consequences?
3. can we prevent a normative effect of the publication?

www.let.ru.nl/corpusngt/

Onno Crasborn: o.crasborn@let.ru.nl

