Danish University Colleges

Compiling a Sign Language Dictionary

Some of the Problems faced by the Sign Language Lexicographer

Kristoffersen, Jette Hedegaard; Troelsgård, Thomas

Compiling a sign language dictionary
Some of the problems faced by the sign language lexicographer

Jette Hedegaard Kristoffersen and Thomas Troelsgård

Centre for Sign Language
University College Capital, Copenhagen, Denmark

As we began working on the Danish Sign Language (DTS) Dictionary, we soon realised the truth in the statement that a lexicographer has to deal with problems within almost any linguistic discipline. Most of these problems come down to establishing simple rules, rules that can easily be applied every time you encounter a specific problem while describing a sign, and that enables the lexicographer to consistently answer questions like "what is the base form of this sign?", "how many meanings does this sign have?", "are these two forms two different meanings of the same polysemous sign – or are they homonyms?" and so on. Very often such questions demand further research and can't be answered sufficiently through a simple standard formula. Therefore lexicographic work often seems like an endless series of compromises.

Another source of compromise arises when you set out to decide which information to include in the dictionary. This depends of course on the target user group, but when you aim at fulfilling the needs of several different user groups, which is what an all-round dictionary must do, you easily risk falling between not two, but several stools. When editing the DTS Dictionary we often face this dilemma, as we see DTS learners and teachers as well as native DTS signers as our target users.

In the following we will focus on four problem areas with particular relevance for the sign language lexicographer:

- Sign representation
- Spoken language equivalents and mouth movements
- Example sentences
- Partial equivalence

Sign representation

A crucial question is how to present sign headwords in the dictionary. For spoken and written languages like English or Danish there are norms for spelling, and although some words have competing variants, most words have a spelling and a base form that all language users agree on. For DTS, and probably for many other sign languages, there is no standardised written form of the signs actually used by the native speakers, and there are no recorded, "authorised" base forms of the signs, although many signs have little or no variation.

Existing sign language dictionaries use different approaches for representing the sign lemmas: video clips, drawings, photos, formal notation (HamNoSys, SignWriting etc.), glosses, numbering, spoken language equivalents, prose descriptions of how the signs are produced.

In Denmark there is no standard notation system adopted by the native signers. Some DTS teachers and learners use SignWriting, but the most common notation form is glosses, spoken language words that reflect the meanings of the signs, and are used as mnemonics for the signs. These are not standardised, but made up along the way, based on the context, which means that one sign can have different glosses in different contexts.

As there obviously is no common denominator, we found that the only secure way to render the headwords in the DTS Dictionary would be through video clips.

This works well in the actual sign entries, but for search result lists, example sentences, cross-references etc., that is in situations, where several signs are shown together, or where sign references appear in a text, it would obviously be quite confusing to use videos.

In the DTS Dictionary we therefore chose to add a sign photograph and a gloss to the entry head, see fig. 1, and to use these to represent signs in the search result list, see fig. 2. When sign representations appears in text, the gloss is used alone, see fig. 3.
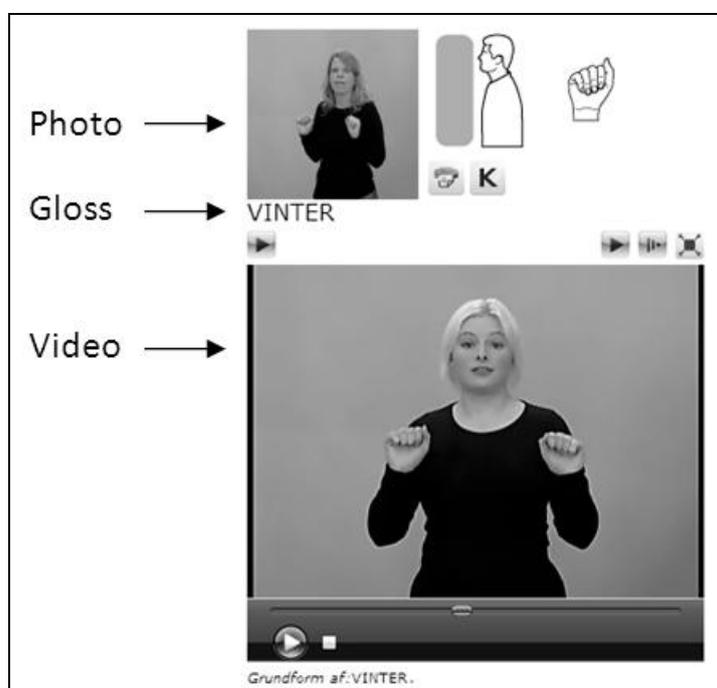

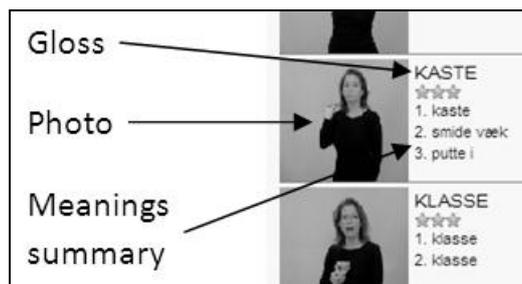**Fig. 1. Sign representation in the entry head of the DTS Dictionary.**

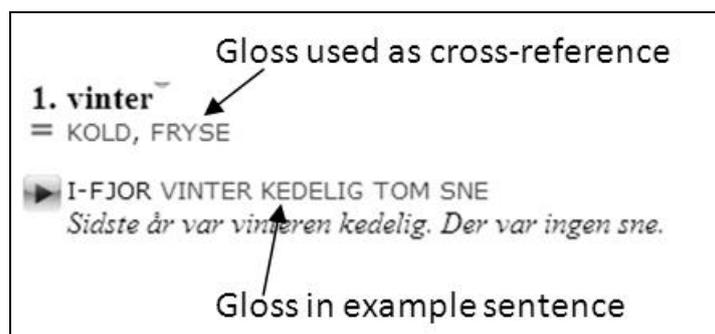**Fig. 2. A sign in the search result list.**

Fig. 3. Glosses used as clickable sign references in the entry.

This decision led to a widespread use of glosses in the dictionary, a solution we would rather have avoided, as it tends to lead the inexperienced dictionary user to the false conclusion, that the sign only (and always) has the meaning expressed by the chosen gloss. On the other hand, we gained the possibility not only to represent the signs by text, but also to easily establish clickable cross-references, as we use unique glosses for the signs.

One problem is how to present the signs uniquely in the dictionary. A related problem is the decision of how the sign should be represented internally in the dictionary database. In the DTS Dictionary signs can be looked up by their manual features, and the search results should be sorted according to these features. For these reasons, a video, a photo and a gloss is obviously not enough – there is also a need for a formal description of the sign. But how detailed should the sign's phonology description be? In the DTS Dictionary we have limited the features by which a sign can be looked up to location and handshape, and likewise decided that these two categories should comprise the key for sorting the search results. Hence, all signs should at least be coded for location and handshape.

As we set up rules for sorting search results, we soon realised that information about location and handshape is not sufficient for sorting the signs satisfactorily. If a combined search on a certain Location and a certain Handshape has several hits, e.g. compare the four signs shown in fig. 4, obviously additional phonological information is needed. For this reason, we decided to include orientation and movement in the phonological description in our database, although this information is not shown explicitly in the user interface.

Fig. 4. Four different signs with coinciding location (palm) and handshape (index finger point).

Spoken languague equivalents and mouth movements

The lemmatisation of the DTS Dictionary is based partly on the signs' manual expression, partly on semantics. A semantic analysis decides which signs we treat as polysemous signs and which signs we treat as homonymous. Every sign meaning has one or more Danish equivalents, if at all possible (otherwise a usage description is provided).

Each equivalent will be marked for mouth pattern. If the Danish word is a mouth pattern that can accompany the sign, it is marked with ⌣ in the entry (fig. 5).
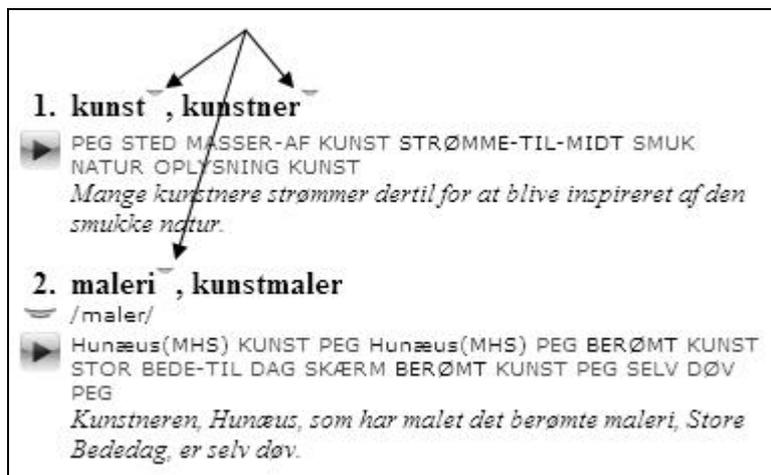
Fig. 5. Mouth pattern markers.

Mouth movements are not solely silent imitations of spoken Danish. DTS has oral components – mouth gestures – not related to the surrounding spoken language (Kristoffersen and Niemelä 2008[1]). Mouth gestures of this kind are also shown in the entry,

---

[1] Kristoffersen, J.H. and Niemelä, J.B. (2008). *How to describe mouth patterns in the Danish Sign Language Dictionary*. In R. Müller de Quadros (ed.). *Sign Languages: spinning and unraveling the past, present and future. TISLR9, forty five papers and three posters from the 9th Theoretical Issues in Sign Language Research Conference. Florianopolis, Brazil, December 2006*. Petropolis: Editora Arara Azul. pp. 230-238.

and additionally a photo of the mouth gesture pops up when its textual representation is moused over (see fig. 6).

Fig. 6. Mouth gestures with pop-up pictures.
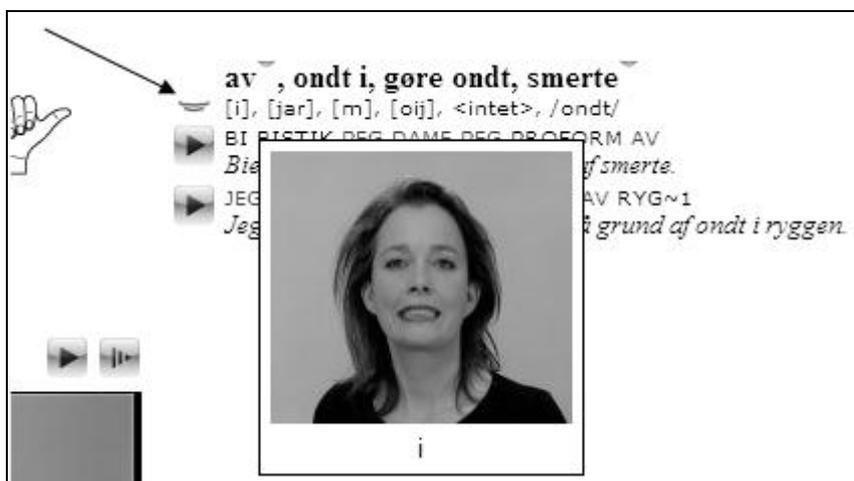
A full inventory of the known DTS mouth gestures is available in the online dictionary.

Example sentences

Every meaning is also illustrated by one or more sentences showing the sign in context. All sentences are represented by a video, by a broad transcription and by a translation into Danish.

The purpose of putting in a broad transcription is to show word order in the sentence and to give the users the possibility to go from one article to another article for a sign used in a sentence

Excerption of sentences

We do not have an annotated DTS corpus. Every sentence was excerpted by hand from a raw corpus of dialogues and monologues – given to us by our group of consultants. The consultants group consists of 0.6% of the deaf population in Denmark, in the age-group 20-60 years. The consultants are recruited from sign language centres across Denmark, typically cities that have hosted deaf schools for many years.

The consultants met with us approximately 4 times a year. Every time we met, we would record a story told by each consultant to another. The topics of the stories were their everyday life, recent experiences or a topic we had chosen. Another task for the consultants was to discuss the use of signs presented by us. All discussions were videotaped. No hearing persons were in the room during the recording. Our raw corpus contains 148 monologues and 75 hours of discussions from these sessions. Discussions were of course used to get information on specific use of signs, but additionally both monologues and discussions were used to elicit signs, analyse sign phonology and semantics, and to build a raw corpus of DTS. From this sample of real language we found

sentences to illustrate the meanings of the signs in the dictionary. The goal for our project has been that every sentence should be taken from real languages samples, but due to restriction on time and money we only managed to let approximately 50% of the sentences originate from real language samples. The remaining half of the sentences was constructed by our deaf lexicographers.

The good example sentence
We didn't just adopt the sentences directly from the source video. We had set up some guidelines on what makes a good sentence, and the sentences were slightly modified to follow these guidelines.

The good context sentence

- reveals something about the core meaning
  A sentence like *This is a cow* does not reveal anything of the meaning of the lexeme 'cow' but the fact that a cow can be. A sentence like *The farmer was milking the cows every morning,* on the other hand, tells us something about a cow belonging to the world of a farmer and being something that you can milk. Equally the fragment *eating strawberries* gives more information on strawberries than *buying strawberries*

- includes a common use of the lexeme
  Sentences serve in this way more than one purpose: firstly the sentence helps to interpret the meaning, secondly the sentence provides the learners with usage examples to copy in their everyday conversation.

- is understandable out of a larger context
  All parts of the sentence should be present. Anaphoric reference may in the real sample only occur as an eye gaze in a direction of a locus, but when we copy only a part of an utterance, this kind of reference must be replaced by a lexical item; a pronoun.

- keeps focus on the sign
  Sentences should be of everyday matters and rather a little dull than too fantastical, in order to keep the user's focus on the matter – understanding the sign – instead of using too much energy trying to understand a seldom experienced situation. To keep focus on the lexemes, the sentences should be showing the sign in a common context, e.g. a sign meaning mixing something should rather be in contexts as mixing water with flour, than in a context mixing raisins with whipped cream. In the last example the user will get an everyday situation to help interpret the meaning of the sign. In the first example the user could be puzzled since – at least in Denmark – mixing raisins with whipped cream to most people is not a well-known everyday activity. So firstly, the sentence will not help understanding the meaning 'mix', secondly, focus will easily be on "what on earth are they cooking here?"

- has a high truth value
  Sentences like: *All fish in Danish lakes have died from lack of oxygen* were modified to *Many fish in Danish lakes have died from lack of oxygen* – again to keep focus on the sign language - lies are disturbing and take away focus.

- must not be offending in any way
  Sexual matters, politics, religious and race-matters should be kept on a minimum.

- does not reveal the source of the sentence
  Our consultants constitute 0, 6% of the deaf population in Denmark. They are all prominent in the deaf community – being an active member of deaf clubs etc. was one of the criteria for being asked to contribute to the project. This means, that our consultants' family, work and life are known to a lot of people. The deaf community is very small in Denmark, and it is essential to us that the sentences in the dictionary not in any way reveal any personal information. Of course all names on persons and places have to be changed, but also other signs, for example the sign for a motorbike has to be changed into a sign for another kind of vehicle, golf has to be changed into another sport etc. to keep the anonymity of the source.

- should contribute to the balance in topics
  At a point we discovered that the sign MOTHER occurred in a lot of sentences, but the sign FATHER only in a few, we had a lot of occurrences of the sign for 'motorbike' but none for 'bike', lots for 'football' and none for 'volleyball' etc. To get a balance in topics we therefore sometimes substitute one noun for another, both belonging to the same semantic field. The user gains from this not only because more topics are represented in the sentences, but also because signs that has not yet been analysed firmly, and has not yet their own entry in the dictionary, can be found in a sentence. Through this substitution possibility we have ensured that a wider range of signs can be watched in the sentences. E.g. we have included all Danish towns as lemmas if the town had more than 20.000 inhabitants. This level exclude a little town that is very prominent in Deaf life in Denmark, so we put in four sentences that include the sign for this town. This way the user can see how to pronounce the sign for this town, even if the town is too small to meet our criteria for having its own entry.

- should not be a definition of the sign
  We distinguish between the definition of a sign and the sign used in real language. In the DTS Dictionary, the meanings of the signs are explained through Danish equivalents, complemented with one or more example sentences. We do not – yet – have definitions of the sign meanings, but we would like to be able to supply definitions later, and we don't want them to clash with the usage examples.

An example
(1) and (2) illustrate how a sentence can be moderated. (1) is the original utterance from our consultant and (2) is how it ended up in the dictionary, when it was turned into a context free sentence with no hesitations from the speaker

 (1) PRON PLACE PF NAME AR[T]/ USE-TO NAME NEW ART SQUEEZE~1/SQUEEZE~2/SQUEEZE~3/ WHAT 1.p WHAT/SOME THINK CAUSE PRON LOTS-OF ART ALSO FLOCK-TO/PRON gesture SEE /LAND[SCAPE]/ BEAUTY LANDSCAPE/ ENLIGHTENMENT/ ART/GO-TO/

(and) that place is called art... use to be called the artist squeeze... squeeze … squeeze, why I don't know, some people think the reason is that also a lot of artists go to this place... erm... to see land... to get inspired by the beautiful landscape, they go there.

(2) PRON PLACE LOTS-OF ART FLOCK-TO/ BEAUTY LANDSCAPE/
ENLIGHTENMENT/ ART/

Lots of artists go to that place to get inspired by the beautiful landscape.

Partial equivalence

Partial equivalence is something that causes a great many misunderstandings for learners of a new language.

The classic example 'tree' - 'wood' often is used to show partial equivalence between languages and it is very useful when comparing Danish and DTS as shown in fig 7.
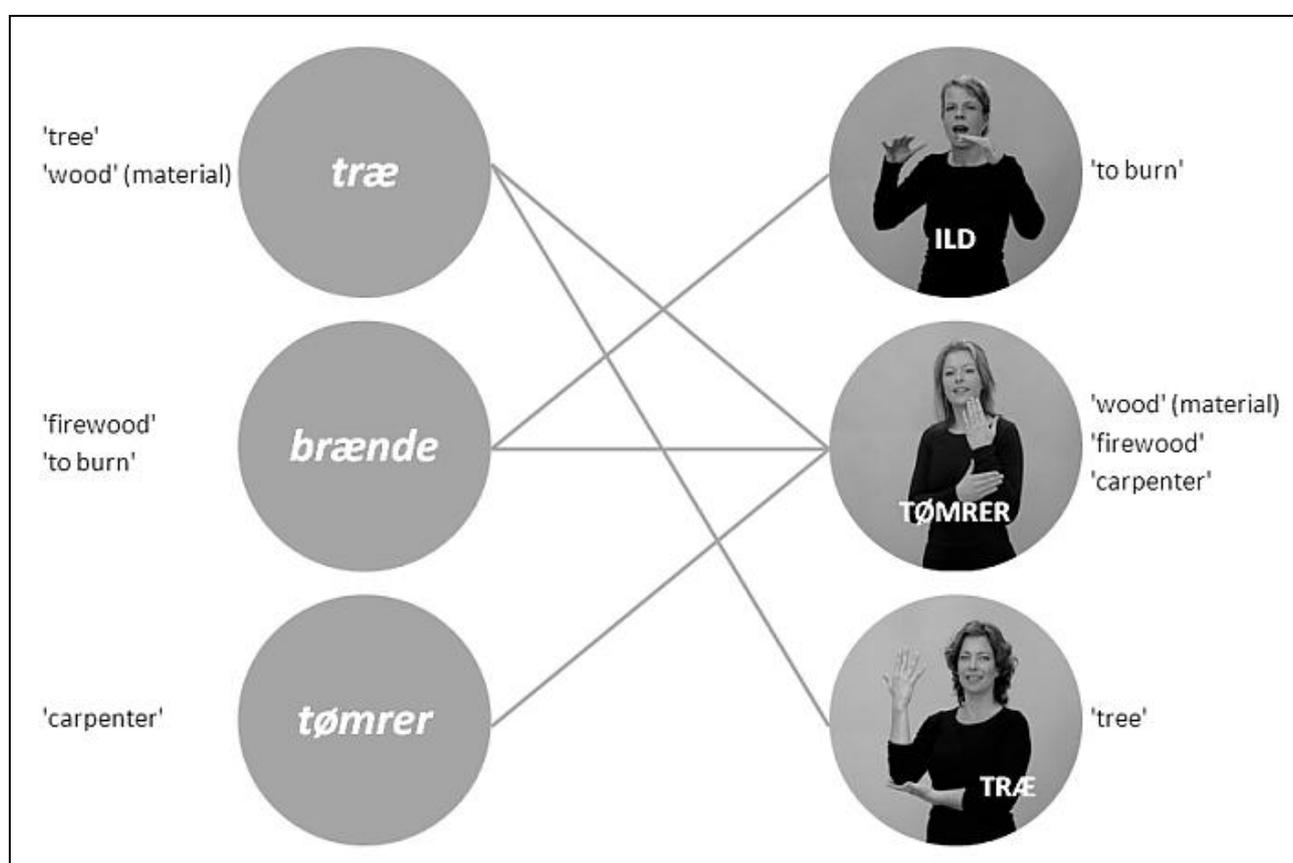

Fig. 7. Partial equivalence Danish – DTS.

The sign in the middle of fig. 7, TØMRER means both 'wood' (the material), 'carpenter', and 'firewood', and it has three different equivalents in Danish.

When the sign has more than one equivalent in spoken Danish, this will be shown to the user in the DTS Dictionary either as several meanings in the entry or as cross-references to partial synonyms, e.g. fig. 8.
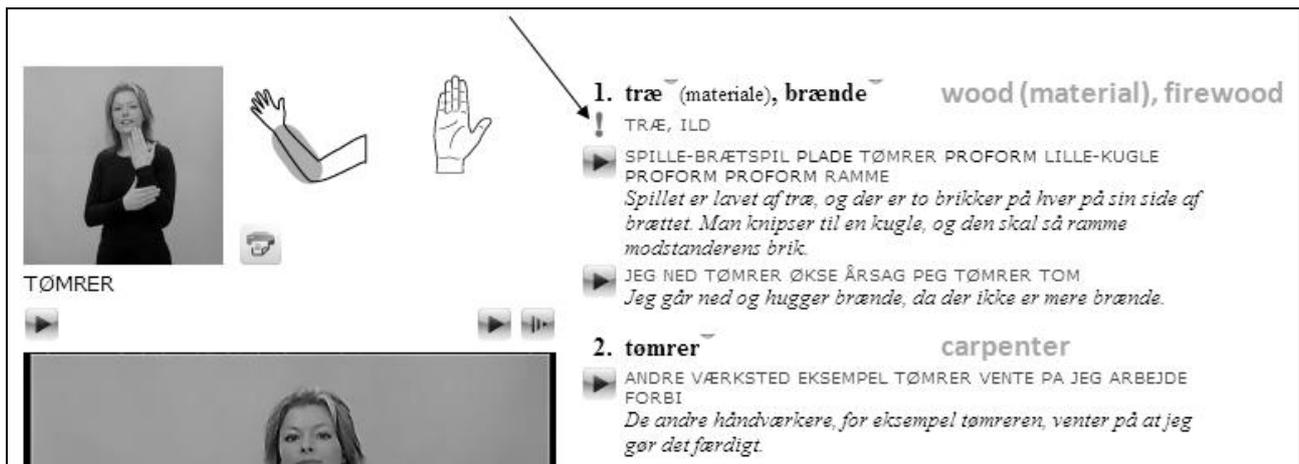
Fig. 8. Entry for TØMRER ('carpenter'), with "polysemous equivalent"-references.

Meaning number one has two partial synonyms as Danish equivalents: *træ* ('wood') and *brænde* ('firewood). Meaning number two has one Danish equivalent: *tømrer* ('carpenter').

The task is not that simple when the partial equivalence goes the other way around:

The Danish word *træ* has two different equivalents in DTS: TØMRER and TRÆ (fig 7.)

For a case like this we use e.g. superordinate concepts to discriminate the meanings: 'wood' (material) – and we also use an exclamation mark as a kind of alert signal to let the user know that he or she could risk choosing the wrong sign due to partial equivalency (fig. 8). The user can then click on the sign next to the exclamation mark and be taken to the sign TRÆ 'wood', the correct translation of the alternative meaning of the polysemous Danish word *træ*.

The other Danish polysemous word *brænde* has two different translations meaning 'firewood' or 'burn'. The other sign next to the exclamation mark ILD (fig. 8) takes the user to the sign meaning 'burn'.

– – –

The problem areas treated above represent some of the particular problems that face a sign language lexicographer. The solutions that we have chosen in the DTS Dictionary often reflect the compromises that we have made in order to achieve our goal: an allround dictionary, that at the same time is comprehensive and easy to use, and which could be of use to a wide range of users – from beginning learners of DTS to native DTS signers.

The DTS Dictionary is freely accessible at www.tegnsprog.dk