This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Salonen, Juhana; Kronqvist, Antti; Jantunen, Tommi

Title: The Corpus of Finnish Sign Language

Year: 2020

Version: Published version

Copyright: © 2020 European Language Resources Association (ELRA)

Rights: CC BY-NC 4.0

Rights url: https://creativecommons.org/licenses/by-nc/4.0/

Please cite the original version:

Salonen, J., Kronqvist, A., & Jantunen, T. (2020). The Corpus of Finnish Sign Language. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), 9th Workshop on the Representation and Processing of Sign Languages : Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives (LREC 2020) (Article 197-202). European Language Resources Distribution Agency. https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/SIGN2020book.pdf

The Corpus of Finnish Sign Language

Juhana Salonen, Antti Kronqvist, Tommi Jantunen

University of Jyväskylä, Sign Language Centre, P.O. Box 35, FI-40014 University of Jyväskylä, Finland
{juhana.k.salonen, antti.r.kronqvist, tommi.j.jantunen}@jyu.fi

Abstract

This paper presents the Corpus of Finnish Sign Language (Corpus FinSL), a structured and annotated collection of Finnish Sign Language (FinSL) videos published in May 2019 in FIN-CLARIN's Language Bank of Finland. The corpus is divided into two subcorpora, one of which comprises elicited narratives and the other conversations. All of the FinSL material has been annotated using ELAN and the lexical database Finnish Signbank. Basic annotation includes ID-glosses and translations into Finnish. The anonymized metadata of Corpus FinSL has been organized in accordance with the IMDI standard. Altogether, Corpus FinSL contains nearly 15 hours of video material from 21 FinSL users. Corpus FinSL has already been exploited in FinSL research and teaching, and it is predicted that in the future it will have a significant positive impact on these fields as well as on the status of the sign language community in Finland.

Keywords: Corpus of Finnish Sign Language, Language Bank of Finland, Finnish Signbank, annotation, metadata, research, teaching

1. Introduction

This paper presents the Corpus of Finnish Sign Language (Corpus FinSL). The corpus was published in FIN-CLARIN's Language Bank of Finland in May 2019 as a result of the four-year (2014–2018) CFINSL (Corpus of Finland´s Sign Languages) project led by the University of Jyväskylä, Finland (see Salonen et al., 2016).[1] The aim of the CFINSL project was to collect, process and make openly available narratives and conversations in FinSL and Finland-Swedish Sign Language (FinSSL), which are the two official sign languages in Finland. During the project, video material was recorded from 91 FinSL users and 12 FinSSL users of different ages from all over Finland.

The published material includes only FinSL data and comprises nearly 15 hours of signing recorded in 2014 from 21 signers. The material has been annotated in ELAN[2] (Max Planck Institute in Nijmegen; Crasborn & Sloetjes, 2008) for signs and Finnish translations. In addition to this, the published material includes anonymized metadata structured according to the IMDI standard (see Section 3). All of the published material is available for academic use under two licenses: a part of the data (elicited narratives) is licensed with Creative Commons 4.0 BY-NC-SA and a part (conversations) with FIN-CLARIN RES (see Table 1 and Section 2). A summary of the content of the published corpus material is shown in Table 1.

All of the material	14 hours and 22 minutes
Elicitated narratives (CC)	5 hours and 4 minutes
Conversations (RES)	9 hours and 18 minutes
Video files	343 mp4-files
Annotation files	142 files (eaf +pfsx)
Number of informants	21 informants

Table 1: Statistics of Corpus FinSL in the Language Bank of Finland.

2. The Subcorpora and Their Motivation

Corpus FinSL in the Language Bank of Finland has been divided into two subcorpora on the basis of licenses and the content of the video material: Elicited narratives and Conversations. The Elicited narratives (Corpus FinSL-elicit) are publicly available for researchers, educators and the general public under the newest non-commercial Creative Commons license (CC 4.0 BY-NC-SA).[3] These narratives consist of signed retellings of cartoon strips, videos and picture books. Access to the dataset of Conversations (Corpus FinSL-conv) is restricted to academic use only; it requires a research plan as well as personal access rights, in accordance with the RES license of the Language Bank of Finland.[4] The material comprises conversations about different topics such as work and hobbies as well as deaf culture. The whole data consists of seven different tasks, which are presented in Table 2. The data of Corpus FinSL-elicit contains tasks 3, 4 and 5 (marked with an asterisk) while the data of Corpus FinSL-conv includes tasks 1, 2, 6 and 7. The structure of Corpus FinSL in the Language Bank of Finland is shown in Figure 1.

Task types:
(1) presenting oneself
(2) telling about one's hobby/work
*(3) signing cartoon strips (Ferd'nand)
*(4) signing a video story (Mr. Bean and Laurel & Hardy videos)
*(5) signing from a picture book (*The Snowman* and *Frog, Where Are You* picture books)
(6) discussing an event related to Deaf culture
(7) free discussion

Table 2: The task types of Corpus FinSL. Tasks marked with an asterisk are freely available whereas tasks without an asterisk are restricted to academic use only.

---

[1] http://r.jyu.fi/tTc
[2] https://tla.mpi.nl/tools/tla-tools/elan/
[3] https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode
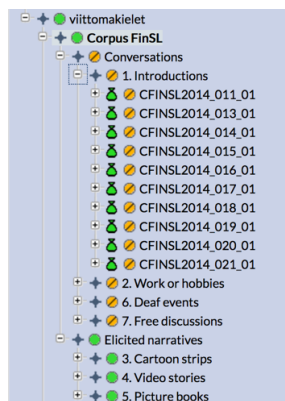[4] https://www.kielipankki.fi/support/clarin-eula/

Figure 1: Screenshot showing the structure of Corpus Fin-SL in the LAT platform of the Language Bank of Finland.

The division of Corpus FinSL material into two subcorpora is primarily due to compliance with the EU Data Protection Regulation (GDPR), which came into force in May 2018. Whereas the subcorpora of Elicited narratives contains only thematically restricted monologues, in which the informants replicate the stories of different picture and video materials, the subcorpora of Conversations includes more thematically free dialogues, in which the informants may also indirectly disclose personal information from third parties. Permission is never granted for the free disclosure of this personal information, so access to the conversation data has been restricted. The decision was made in collaboration with the legal department of the University of Jyväskylä and FIN-CLARIN.

In general, the informants' consent was collected in two phases. First, after a filmed session (in 2014), the informants were asked to choose either the Yes or No option for each of the following five statements:

1. Video material can be used for research purposes, but publishing video clips or still images is prohibited
2. Video material can be presented in public events (e.g. academic presentations and teaching)
3. Still images can be taken from the video material for publications (electronic or paper)
4. All the video material can be published electronically e.g. in the Internet
5. The name of the participant can be mentioned in publications.

When the publication of the material became imminent (in 2018), after the GDPR had come into force, it was necessary to update the consents originally collected with additional information. While the original consent treated the online publication of the material as one general entity, the additional consent form that we devised required the informants' explicit permission for the long-term preservation of the material in the Language Bank of Finland and the free publication of each communication assignment (seven task types of varying nature and privacy) on the same platform. Pursuant to the GDPR´s

guidelines, the informants were also asked to explicitly consent to the inclusion of personal data belonging to certain specific categories. In particular, this concerned the information they had given us with respect to their hearing status, which falls into the category of information concerning special groups mentioned in the GDPR. This type of data was included not only in the background forms collected from the informants but also, for example, arises in interactive task 1, where the informants naturally identify themselves as deaf when presenting themselves in a way specific to the sign language culture.

In the additional consent form, the informants were asked to choose, as in the original consent form, either the Yes or No option for questions 1, 2 and 4. The third question, on the other hand, included a list of all seven task types, of which the informants then selected those which they allowed to be published. Before completing the consent form, the informants received all of the necessary information in both Finnish and FinSL.

1. In addition to my prior consent (dated X) in the attachment, I consent to the licensing of the recording that was the subject of this consent, under a *Creative Commons-ByAttribution-NonCommercial-ShareAlike 4.0 International* (BY NC SA 4.0) license, and agree that the University of Jyväskylä is the copyright holder of this recording
2. In addition, I consent to the transfer for long-term preservation of the licensed recording that was the subject of this consent to FIN-CLARIN's Language Bank of Finland, where it may be used for research purposes on the basis of appropriate research plans
3. I also consent to the publication of the licensed recording that was the subject of this consent in the Language Bank for the tasks that I have ticked below, meaning that anyone can see and use them
4. I consent to the inclusion in the recording of information that falls into specific categories of personal data, such as the degree of my hearing

In practice, the additional consent form also required informants to consent to the licensing of their entire material under the Creative Commons 4.0 BY NC SA license, which ultimately prohibits commercial use. However, as explained above, only the Elicited narratives subcorpora is covered by this license. Free publication of the Conversation subcorpora is possible in the future if the annotations and video material comprising it are anonymized for any third party related information.

3. Annotations and Metadata

The published Corpus FinSL contains over 107,000 glossed sign tokens. The annotation process (Sections 3.1 and 3.2) began with the identification of sign units, the distinction of their meanings and forms (see Salonen et al., 2016) as well as provisional information about the grammatical behavior of the sign (e.g. negation), and the

translation of chunks of signed utterances into Finnish. The metadata (3.3) describes the generic nature of the dataset, the individuals present in the collection, the content of the videos, and the formats of the video and annotation files.

3.1 Sign Level Annotation

Unity and consistency are the prerequisites for building a functioning corpus. This means that common principles and annotation guidelines must be agreed between all annotators. Annotation conventions[5] have been developed in several sign language corpus projects (e.g. Johnston, 2016 Australia; Schembri et al., 2013 the United Kingdom; Crasborn et al., 2015 the Netherlands; Wallin & Mesch, 2018 Sweden). In the CFINSL project, we began to develop the conventions during the basic annotation (see Keränen et al., 2016). The first version of the annotation conventions was released in spring 2018 and included the guidelines for sign level annotation in the CFINSL project (see Salonen et al., 2018 as well as Tables 3 and 4). There are guidelines in the conventions for the recording of different parts of the lexicon, such as lexical signs of varying degrees (see Jantunen, 2018). Guidelines related to creating sentence level translations were added to the second version of the conventions, published in February 2019 (Salonen et al., 2019).

Category	Example
Lexical signs	a common/distinct ID-gloss:
o Phonetic variants (1-2 different parameters)	FATHER(Ax), FATHER(G) => FATHER
o Lexical variants (2-3 different parameters)	with e.g. a handshape code: DO(BB) vs. DO(SS)
o Polysemic signs	BALL, WORLD => BALL
o Homonym signs	BEACH, BORDER => BORDER
Numeral signs (_num)	SIX-YEAR_num, ONE-WEEK-EARLIER_num
Pointings (OS:)	OS:, OS:me
Depicting signs (_kv + a subclass code)	_kvkk (a whole entity classifier)
Gestural signs (_ele)	PALM-UP_ele
Fingerspellings (_sa)	t-o-m-i_sa

Table 3: Examples of the conventions of sign level annotation.

The finalized sign-level annotation of Corpus FinSL material exploits ID-glosses, which means that signs of the same form (homonymous, polysemic, and phonetically

variable) are assembled under the same tags (for the annotation process more generally, see Salonen et al. 2016). In practice, an ID-gloss refers to a label selected to represent sign tokens that have a similar form but whose meanings may vary in a corpus (Johnston, 2008, 2010). For example, in FinSL there is a manually articulated sign, the same form of which can mean 'everyday', 'jeans', 'rural', 'fresh' or 'orange', depending on the context in which it appears. Instead of tagging the tokens with a context-specific meaning gloss, the sign is glossed with a single label EVERYDAY to represent all tokens of the same form. Thus the ID-gloss does not indicate a meaning translation of the sign, but rather functions as an identifier agreed upon by the annotators. ID-glossing allows one to search the data more efficiently than glossing according to a contextual meaning. The basic annotation of Corpus FinSL has been focused on identifying sign tokens as much as possible in a systematic and fast manner from the perspective of the annotators without actual research guidelines.

Type of grammaticality	Code
Negation	@neg
Repetition+plural	@toisto
Compound sign	@y
List buoy	@poijul
Lexicalized fingerspellings	@sv
Foreign borrowings	@lv

Table 4: Examples of the conventions for the grammatical behavior of the sign.

We implemented ID-glossing on two interconnected platforms: glosses that are temporarily connected to video material in ELAN are also connected to a database of Finnish Signbank[6] via a network connection. Finnish Signbank is the lexicon database built for FinSL and FinSSL; its basic function is to serve as a tool for annotating sign language texts. In addition to the gloss, Finnish Signbank contains the citation form of the sign on video(s), the sign´s Finnish or Swedish equivalents, and any further information on the sign, if necessary. The database can be updated as annotation work progresses. Figure 2 contains four videos of the same sign form but with different meanings. The difference in meaning can be detected, for example, with the help of a mouthing.

The ELAN program includes a feature that allows the program to access external controlled vocabulary (ECV) maintained by an external web server when annotating. In the CFINSL project, a controlled vocabulary (CV) was placed in Finnish Signbank by developing the Signbank platform for this purpose. The CV allows the annotator to

---

check whether the gloss is already in the database when labeling the annotation cell. If the gloss already exists, the annotator can select it from the list, thus avoiding spelling mistakes, which are easily generated in non-automated transcription. If the gloss has not yet been created for that sign, one can add it to the database and create a new gloss record with videos, translations, and more. Figure 3 illustrates how the ELAN program makes it possible, when creating annotation cell content, to search for a suitable gloss from Finnish Signbank with the help of either the ID-gloss (left-hand column of the box) or its translation equivalents (right-hand column of the box). If necessary, it is also possible to check the video(s) on the gloss page of Signbank. (Salonen et al., 2018.) Manually executed gloss and translation changes in Signbank are automatically updated on all linked annotation cells with a continuous ECV connection.
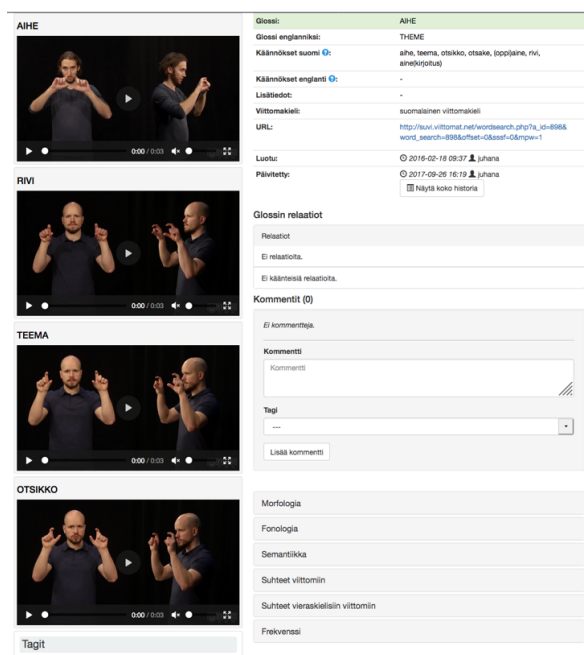

Figure 2: A view of a gloss page in Finnish Signbank.

Finnish Signbank has been developed by the CFINSL project, which has been cooperating with the corpus and dictionary work of the Finnish Association of the Deaf. The database is based on Auslan Signbank,[7] originally developed for Australian Sign Language, and its subsequent application, the Dutch Signbank database. Source codes from all Signbank databases are available on the Github version control site,[8] where the database structure and features of Finnish Signbank can also be found, documented in a user-friendly way (named FinSL-signbank wiki). Signbank development started some ten years ago with international collaboration between research teams in Australia, the Netherlands, Finland and the United Kingdom (Cassidy et al., 2018). Finnish Signbank

comprises two different and independent lexicons. The lexicon of Corpus FinSL contains all of the fixed signs (i.e. lexical signs), as well as a small group of depicting, gestural (emblems) and pointing signs (for further analysis) that have been found in the corpus material. In addition, there is the lexicon of the Kipo corpus of the Finnish Association of the Deaf, which is based on annotated material (approximately 2.5 hours) from the Language Policy Programme for the National Sign Languages in Finland (Kuurojen Liitto ry, 2015).
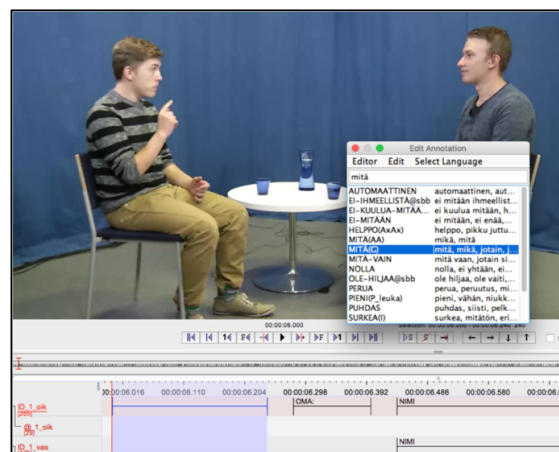

Figure 3: A view of annotation in ELAN using a controlled vocabulary hosted by Signbank.

3.2 Sentence Level Annotation

The video material has also been annotated on the sentence level. In practice, this means that the signing has been translated into Finnish. At the beginning of the translation, meaningful sentences were separated from the referenced text stream by the translators' intuition, without further distinction of sentences, which is the task of the actual study after basic annotation. The translation is in a form that takes into account the way in which the source language is expressed, both manually and non-manually (with the head, body and face).

In addition, to the translations have been added parts in parentheses which a fluent Finnish text requires, but which are not made visible in the preceding discourse context or which may not be required at all in sign language text (e.g. the subject of the sentence, a copula, some conjunctions; see Example 1). The translation guidelines are described in more detail in the annotation conventions (see Salonen et al., 2019).

(1)   LOOK-AT   OUTSIDE   RAIN   SNOW   RAIN
        (He/she) notices, (that) it is snowing outside.

The translation provides a more complete view of the signed texts, as ID-glossing focuses solely on manual articulation. The translation can also be used to check what

---

meaning the ID-gloss is referring to. The translators have also made Finnish translations of signs for gloss pages in the Finnish Signbank database, according to the meanings of the signs marked with ID-glosses in connection with the translation.

3.3   Metadata

The anonymized metadata of Corpus FinSL transferred to the Language Bank of Finland are described in accordance with the IMDI (*ISLE Meta Data Initiative*) standard. IMDI is a description standard developed at the Max Planck Institute in Nijmegen for consistency in the description of multimedia and multimodal language materials.[9] In accordance with IMDI standards, the CFINSL project produced a general description of the material (*Corpus FinSL*), the underlying project (*CFINSL Project*), and the contents of both of the subcorpora (*Elicited narratives; Conversations*) as well as their communication tasks (1-7). In addition, a session-specific description was made of the individual communication tasks of each pair (*Session*), which included information about the participants (*Actors*); the quality of the communication situation, its interactivity and its collection method (*Content*); video materials (*MediaFiles*); and annotations (*WrittenResources*).

Background information about the informants was collected very extensively during the CFINSL project collection. However, the IMDI description built into the Language Bank eventually selected only an anonymous code that identifies the individual, age group, gender, area of residence, and handedness (left/right).

4.   Exploitation of Corpus FinSL in Research and Teaching

The material of the CFINSL project in general and of Corpus FinSL in particular has already been exploited in several research projects focusing on FinSL. First of all, the corpus provided new insights into the study of FinSL word order (e.g. Jantunen 2017), lexicon (Takkinen et al., 2018) and nonmanuality (Puupponen 2018). Currently, the corpus is the main source of data for a larger project which is investigating the role of gesturality in language by focusing on the use and variation of constructed action in FinSL. In addition, the material has been used in comparative studies of sign languages (Jantunen et al., 2016; Puupponen et al., 2016) and in Master's theses completed on the subject of Finnish Sign Language (e.g. Syrjälä, 2018; Puhto, 2018). In general, the existence of the corpus has already had a significant impact on the research tradition of FinSL by requiring that individual studies should be more closely connected to the material that has been collected.

Another, but globally not so recognized, area where Corpus FinSL has made a contribution in the field of FinSL is teaching. Corpus FinSL was taken into account in the planning of the new curriculum for FinSL in Jyväskylä University (2020-2023): the corpus material is included in the course descriptions and targeted as learning outcomes

of different courses. This obliges teachers to apply the corpus in teaching.

In addition, in the fall of 2019, in-service training[10] for sign language teachers funded by the Ministry of Education and Culture was started at the Sign Language Centre, University of Jyväskylä. The aim of the in-service training is to keep sign language teachers informed about new research. In-service training includes three different courses: FinSL grammar, Deaf Studies, and (sign) language acquisition and language assessment. We have made new learning materials, especially in the FinSL grammar course, which is based on Corpus FinSL data. At the moment, most sign language teaching materials are not aligned with the newest research findings, especially within corpus-based research.

The learning materials we have created are given to participants on the in-service training course so that they can exploit the materials in their own teaching. We plan in the future to publish the learning materials on a website where they will be freely available to everyone for teaching purposes.

From the experience we have gathered so far of using the corpus in teaching we have seen that the corpus can be used in teaching in at least three different ways.

1. Examining the corpus data; finding and discussing relevant language-related sign units and language structures.
2. Annotation of corpus data by students. This helps to internalize language usage and variation in language units.
3. Searching for corpus data to illustrate a theoretical perspective (e.g. sign types, word order in FinSL); additional annotation may be required.

All this requires appropriate (research) literature; corpus data is used and discussed in relation to the (research) literature. In addition, corpus data can be used to teach deaf culture or sign language communication. The data contains e.g. conversations about events related to deaf culture, and offers interesting topics for discussion. The corpus includes interaction between two persons and therefore also offers an opportunity to analyze various interaction practices.

5.   Discussion and Conclusion

In this paper we have described the content of Corpus FinSL as it is published in the Language Bank of Finland. We have presented the annotation guidelines, metadata and exploitation of the corpus material in research and teaching. The extensive electronic and computer-readable material offers new opportunities for quantitative and qualitative research on the sign languages used in Finland; it already does this for FinSL, and later it will do the same for FinSSL. The wide-ranging, multi-person material can be used to examine, for example, the signing of native signers of different ages from around Finland, as well as

---

differences in signing and language structure between different types of text genres. The extensive, partly publicly available data also allows for a completely new way of comparing sign languages. This is supported in particular by the use of similar data collection methods in corpus projects in different sign languages.

Corpus FinSL will have a significant impact on Finland´s sign language community and the social status of sign languages. For sign language people, it provides an opportunity to develop language awareness of their mother tongue, which many have not been taught in basic education. For those using sign language as a foreign language – such as sign language interpreters – the corpus provides educational material on, for example, recognizing the socio-linguistic differences between language users. The corpus material will also continue to be used in an ongoing project at the University of Jyväskylä to develop in-service training for sign language teachers in Finland. Finally, in addition to its teaching and training applications, the corpus may in the future serve as a tool for developing language management and language planning.

6. Acknowledgements

7. References

Cassidy, S., Crasborn, O., Nieminen, H., Stoop, W., Hulsbosch, M., Even, S., Komen, E. & Johnston, T. (2018). Signbank: Software to Support Web Based Dictionaries of Sign Language. *Proceedings - The 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, 2359-2364.

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. *Proceedings - The 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, 39-43.

Crasborn, O., Bank, R., Zwitserlood, I., Kooij, E., Meijer, A. & Sáfár, A. (2015). Annotation Conventions for The Corpus NGT. Version 3. Radboud University Nijmegen: Centre for Language Studies & Department of Linguistics.

Jantunen, T. (2018). Viittomakielet hybridisysteemeinä: hämärärajaisuus ja epäkonventionaalisuus osana viittomakielten rakennetta. *Puhe ja Kieli* 38(3), 109-126.

Jantunen, T. (2017). Fixed and NOT free: Revisiting the order of the main clausal constituents in Finnish Sign Language from a corpus perspective. *SKY Journal of Linguistics* 30, 137-149.

Jantunen, T.; Mesch, J.; Puupponen, A. & Laaksonen, J. (2016). On the rhythm of head movements in Finnish and Swedish Sign Language sentences. *Proceedings - Speech Prosody 2016*, 850-853.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. *Proceedings - The 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, 82-87.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics* 15 (1), 106-131.

Johnston, T. (2016). *Auslan Corpus Annotation Guidelines*. Centre for Language Sciences, Department of Linguistics, Macquarie University (Sydney) and La Trobe University (Melbourne), Australia.

Keränen, J., Syrjälä, H., Salonen, J. & Takkinen, R. (2016). The Usability of the Annotation. *Proceedings - The 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining,* 111-116.

Kuurojen Liitto ry (2015). *Suomen viittomakielten kielipoliittinen ohjelma 2010 -korpus, annotoitu versio.* Kielipankki. http://urn.fi/urn:nbn:fi:lb-2014073031

Puhto, J. (2018). *Päänpudistuksen käyttötavat ja frekvenssit suomalaisessa viittomakielessä.* Suomalaisen viittomakielen pro gradu -tutkielma. Kieli- ja viestintätieteiden laitos, Jyväskylän yliopisto.

Puupponen, A., Jantunen, T. & Mesch, J. (2016). The Alignment of Head Nods with Syntactic Units in Finnish Sign Language and Swedish Sign Language. *Proceedings - Speech Prosody 2016*, 168-172.

Puupponen, A. (2018). The relationship between the movements and positions of the head and the torso in Finnish Sign Language. *Sign Language Studies* 18(2), 175-214.

Salonen, J., Takkinen, R., Puupponen, A., Nieminen, H. & Pippuri, O. (2016). Creating Corpora of Finland´s Sign Languages. *Proceedings - The 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, 179-184.

Salonen, J., Wainio, T., Kronqvist, A. & Keränen, J. (2018). *Suomen viittomakielten korpusprojektin (CFINSL) annotointiohjeet.* 1. versio. Kieli- ja viestintätieteiden laitos, Jyväskylän yliopisto. http://r.jyu.fi/ygQ

Salonen, J., Wainio, T., Kronqvist, A. & Keränen, J. (2019). *Suomen viittomakielten korpusprojektin (CFINSL) annotointiohjeet.* 2. versio. Kieli- ja viestintätieteiden laitos, Jyväskylän yliopisto. http://r.jyu.fi/ygR

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S. & Cormier, K. (2013). *Building the British Sign Language Corpus*. Language Documentation and Conservation 7, 136-154.

Syrjälä, H. (2018). *Hakukysymysviittoman paikka suomalaisessa viittomakielessä.* Suomalaisen viittomakielen pro gradu -tutkielma. Kieli- ja viestintätieteiden laitos, Jyväskylän yliopisto.

Takkinen, R., Keränen, J. & Salonen, J. (2018). Depicting Signs and Different Text Genres: Preliminary Observations in the Corpus of Finnish Sign Language. *Proceedings - The 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, 189-194.

Wallin, L. & Mesch, J. (2018). Annoteringskonventioner för teckenspråkstexter. Version 7. Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet.