

SUOMEN VIITTOMAKIELTEN KORPUSPROJEKTIN (CFINSL) ANNOTIOINTIOHJEET



JYVÄSKYLÄN YLIOPISTO

Juhana Salonen, Tuija Wainio,
Antti Kronqvist ja Jarkko
Keränen
Kieli- ja viestintätieteiden
laitos, viittomakielen keskus,
Helmikuu 2019

SISÄLLYSLUETTELO

1 Johdanto	3
2 Annotaatiosolun pituuden määrittely.....	6
3 Leksikotietokanta Signbank	8
4 Viittomien luokittelu ID-glossirivillä.....	11
4.1 Kiinteät viittomat	11
4.1.1 Muita kiinteiden viittomien merkintäkonventioita.....	13
4.2 Numeraaliviittomat	14
4.3 Kuvailevat viittomat.....	15
4.4 Eleet.....	20
4.5 Sormiaakkoset	20
5 Viittoman lisätiedot.....	22
5.1 Rakenneyksikkötietojen merkitseminen ID-glossiin.....	22
5.2 Kieliopillisten piirteiden merkitseminen tytäriville	25
5.2.1 Kielto	25
5.2.2 Monikko ja toisto	27
5.2.3 Listapojut.....	28
5.2.4 Yhdysviittomat	29
5.2.5 Sormiaakkosia sisältävät viittomat.....	32
5.2.6 Lainaviittomat.....	32
5.2.7 Epäselvät viittomat.....	33
5.2.8 Käden jähmettyminen	33
5.2.9 Pään nyökkäys ja pudistus	34
6 Käännöksen periaatteita	35
Lähteet.....	40

1 Johdanto

Jyväskylän yliopiston viittomakielen keskuksessa käynnistyi virallisesti keväällä 2014 neljän vuoden (2014–2018) mittainen Suomen viittomakielten korpusprojekti (CFINSL), jonka aikana on tallennettu Suomessa käytettävää suomalaista ja suomenruotsalaista viittomakieltä. Korpuksella tarkoitetaan sähköisessä muodossa olevaa kirjoitetun, puhutun tai viitotun kielen editoitua ja annotoitua kokoelmaa, jonka avulla voidaan tehdä kielen sanaston, kieliopillisten rakenteiden ja käytön tutkimusta. Annotointi viittaa kirjoitetun, puhutun tai viitotun aineiston kuvaamiseen, luokitteluun ja jäsentelyyn systemaattisella tavalla. Korpusprojektin tavoitteena oli myös käynnistää kuvatun aineiston perusannotointi ja laatia annotointiohjeet.

Kuvaukset valmistuivat syksyllä 2017, jolloin oli kuvattu 91 suomalaista viittomakieltä ja 12 suomenruotsalaista viittomakieltä äidinkielenään käyttävän, eri puolilla Suomea asuvan henkilön viittomista. Kuvaukset järjestettiin pääosin Jyväskylän yliopiston AV-studiolla siten, että viittomakielisiä pyydettiin pareittain kuvattavaksi. Kuvauksia tehtiin myös Helsingissä ja Oulussa. Kuvausten ohella kerättiin taustatietoja viittojista (mm. ikä, sukupuoli, asuinpaikka). Kuvausten ja editoinnin jälkeen korpusaineistoa ryhdyttiin annotoimaan ELAN-ohjelmalla (Eudico Linguistic Annotator), joka mahdollistaa aineistosta tehtyjen merkintöjen ajallisen yhdistämisen videomateriaaliin ja lukuisia erilaisia hakutoimintoja. Korpusaineisto taltioidaan Fin-Clarinin hallinnoimaan Kielipankkiin, josta käsin aineistoa voidaan hyödyntää tutkimus- ja opetustarkoituksiin informanttien antaminen lupien rajoissa. (Salonen ym. 2016.)

Tässä projektissa aloitettiin perusannotointi, jossa on keskitytty lähinnä viittomien erotteluun ja merkitsemiseen. Viittomakielten viittomien nimeämisessä käytetään yleisesti glosseja, jotka ovat kirjoitetun kielen sanoja. Glossiksi valitaan yleensä sana, joka lähimmin vastaa viittoman merkitystä (Johnston 2010, 2016). Toistaiseksi viitotuille kielille ei ole luotu yleisesti käytössä olevaa kirjoitusjärjestelmää. Tämän vuoksi suomalaisen viittomakielen kohdalla käytetään yleensä suomenkielistä glossia, joka kirjoitetaan suuraakkosin ja perusmuotoisena.

Annotointia voidaan toteuttaa monin tavoin. Useissa viittomakielten korpusprojektien annotoinnissa on aluksi keskitytty ID-glossin eli tunnisteglossien merkitsemiseen. Glossi kirjoitetaan sellaisessa muodossa, joka olisi mahdollisimman yksinkertainen haun näkökulmasta (Johnston 2010, 2016). Nimenomaan tietokoneella suoritettavat

hakutoiminnot ovat edellytys sille, että voidaan puhua korpuksesta. Korpusta laadittaessa on olennaista yhtenäisyys ja johdonmukaisuus, jolloin tulee sopia yhteisistä periaatteista ja muodostaa konventioita annotointityölle. Tässä ohjeistuksessa perehdytään CFINSL-korpusprojektin annotointikonventioihin, joissa käytetään ID-glosseja.

Annotointikonventiot perustuvat hakutoimintojen kannalta annotointien systemaattisuuteen ja johdonmukaisuuteen. On siis sovittava yhteiset periaatteet, joita kaikki annotoijat noudattavat. (Keränen ym. 2016.) Esimerkiksi viittomakielessä on viittomia, joilla voi olla useita merkityksiä, kuten suomalaisen viittomakielen TEHDÄ-viittoma. Kyseinen viittoma tuotetaan B-käsimuodolla ja siinä on edestakainen vuoroliike. Viittoma voi tarkoittaa tekemisen ohella työtä, työskentelyä, touhuamista, puuhaa jne. Tällaisessa tapauksessa kaikki saman muotoiset viittomat, huolimatta kulloisesta merkityksestä, merkitään samalla ID-glossilla, jotta ne saadaan yhdellä hakutoiminnolla esiin. Tässä projektissa tehtävää annotointia varten ELAN-ohjelmaan nimettiin 20 riviä seuraavasti (Kuvio 1, Keränen ym. 2018):

ID_1_oik [0]	ID_2_oik [0]
@_1_oik [0]	@_2_oik [0]
ID_1_vas [0]	ID_2_vas [0]
@_1_vas [0]	@_2_vas [0]
ID_huomioita_1 [0]	ID_huomioita_2 [0]
Merkitys_1_oik [0]	Merkitys_2_oik [0]
Merkitys_1_vas [0]	Merkitys_2_vas [0]
Merkitys_huomi [0]	Merkitys_huomi [0]
Käännös_1 [0]	Käännös_2 [0]
Käännöshuomio [0]	Käännöshuomio [0]

Kuvio 1. Annotointirivit.

Punaisella tekstillä nimetyt rivit viittaavat videolla näkyvän parin vasemmanpuoleiseen viittojaan ja sinisellä tekstillä nimetyt rivit oikeanpuoleiseen viittojaan. Kummallekin viittojalle on käytössä 8 emoriviä ja kaksi tytärriiviä.

ID-glossausta varten on oma rivi viittojan molemmille käsille (esim. ID_oik) ja kummallakin rivillä on tytärriivi (esim. @_1_oik) lisämerkintöjä varten. Tytärriivit ovat aina yhteydessä emoriveihinsä. Lisäksi on omat rivinsä kummallekin kädelle merkitysglossausta varten, sekä viittojakohtaiset rivit glossaukseen liittyviä huomioita, käännöksiä ja käännöksiin liittyviä huomioita varten. Merkitys-riveille tulevia annotointeja ei tässä dokumentissa käsitellä, vaan niistä tulee myöhemmin erillinen ohjeistus. Annotointiriveistä ja muista ELANIin liittyvistä ohjeistuksista kerrotaan tarkemmin Keräsen (emt.) ELAN-ohjeissa.

Tässä dokumentissa esitellään ne annotointiohjeet, joiden pohjalta CFINSL-tiimi on annotoinut tähän mennessä. Ohjeet ovat syntyneet annotoijien (Jarkko Keränen, Antti Kronqvist, Juhana Salonen, Anu Savolainen ja Henna Syrjälä) sekä korpustiimissä mukana olleiden (Olga Green, Tommi Jantunen, Henri Nieminen, Outi Pippuri, Anna Puupponen, Silja Ruonala, Ritva Takkinen ja Tuija Wainio) yhteiskeskustelujen tuloksena. Lisäksi kiitos Kuurojen Liiton korpus- ja sanakirjatyön tiimille yhteistyöstä. Annotointiohjeet alkavat annotointisolun pituuden määrittelyllä ja Signbank-tietokantaan tutustumisella, minkä jälkeen keskitytään viittomien alustavaan luokitteluun ja erilaisten lisätietojen merkintään. Ohjeiden lopussa esitellään käännöksen periaatteita, joiden turvin tehdään käännöstä annotoinnin ohella.

2 Annotaatiosolun pituuden määrittely

Annotoinnissa tulee tehdä päätös siitä, miten määritellään viittoman alku ja loppu. Olemme päätyneet Jantusen (2013, 2015) esittämään käsitykseen viittoman pituudesta, joka huomioi usein siirtymäliikkeiksi erotellut viittoman osat osana viittomaa (kuvio 2). Jantusen (emt.) mukaan suurin osa viittoman rakennepiirteistä on muodostettu jo hyvissä ajoin ennen ydinosan alkua (viittoman valmistelu) ja suurin osa niistä voidaan havaita myös ydinosan päättymisen jälkeen (viittoman palautus). Viitotussa tekstissä olevien viittomien yksiselitteinen segmentointi ei välttämättä ole mahdollista, koska viittomat ovat hämärärajoja ja ne esiintyvät sulautuneina toisiinsa kuten puhutun kielen sanat. On tarpeellista erottaa käsitteellisesti toisistaan **viittoma** ja **annotaatiosolu**. Annotaatiosolu on viittoman ”sijainnin” karkea merkitsin, jonka täsmällisen alku- ja loppukohdan määrittäminen ei ole annotoinnin kannalta tärkeää.

Lyhyt



Pitkä

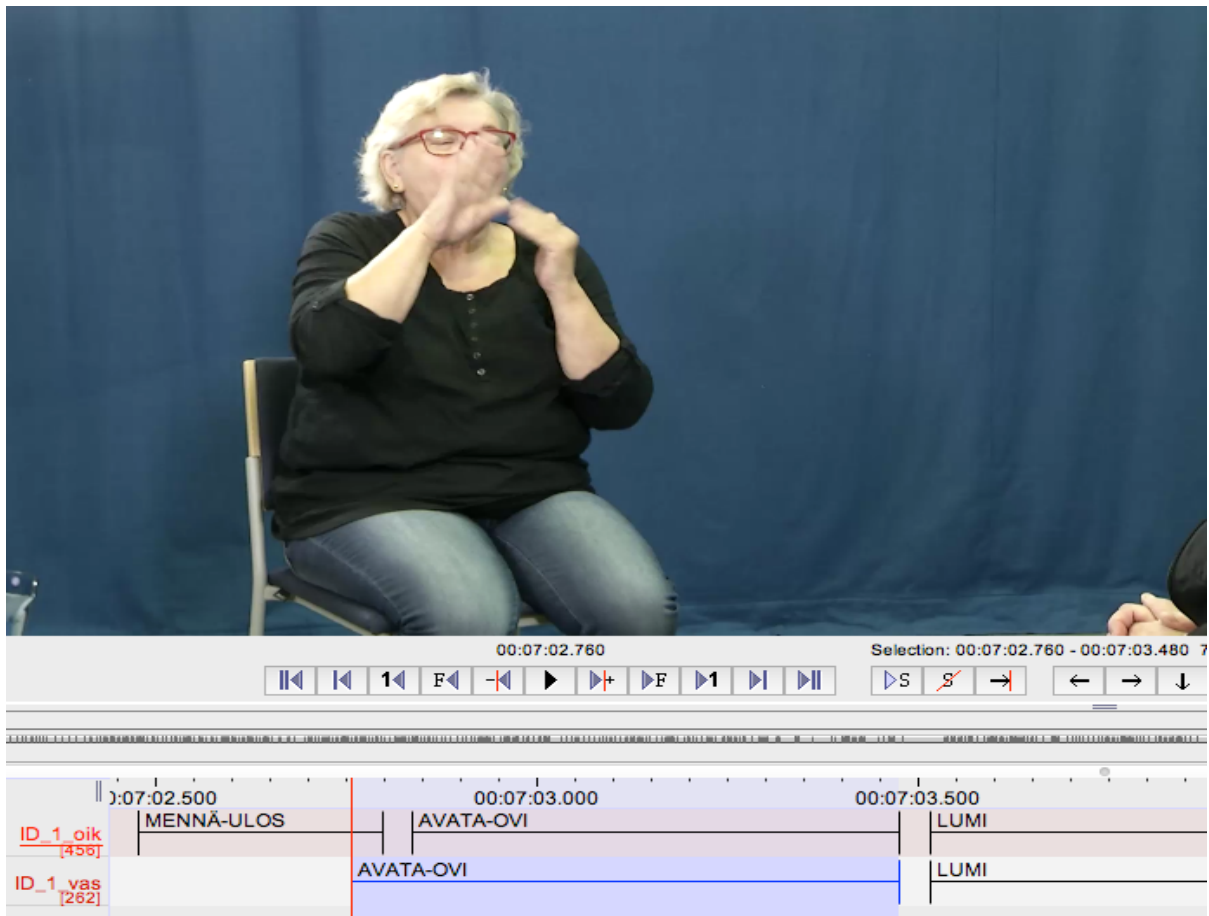


Kuvio 2. Mikä on viittoman pituus? (Jantunen 2014).

Annotaatiosolun pituus määrittyy Jantusta (2013, 2015) soveltaen seuraavasti:

- **Alkaa** videoruudusta, jossa jokin viittoman rakennepiirteistä (yleensä käsimuoto tai orientaatio) on ensimmäisen kerran tunnistettavasti muodostettu.
- **Loppuu** videoruutuun, jossa jokin viittoman rakennepiirteistä (yleensä käsimuoto tai orientaatio) on viimeisen kerran tunnistettavasti läsnä.

- ELANissa annotaatiosolujen väliin on syytä jättää (hakuteknisistä syistä) vähintään **yksi videoruutu (frame)**.
- **Huom!** Molemmat kädet annotoidaan omille riveilleen. Kädet voivat tuottaa itsenäisesti eri viittomia tai kaksikäteisissä viittomissa kunkin käden annotaatiosolun pituus voi vaihdella. Esimerkiksi kaksikäteisessä viittomassa AVATA-OVI (kuvio 3) on vasemman käden viittoman muodostus alkanut ennen oikeaa kättä:



Kuvio 3. AVATA-OVI -viittoman annotaatiosolut ELAN-ohjelmassa.

3 Leksikotietokanta Signbank

Käytämme annotoinnissa ELAN-ohjelman lisäksi leksikotietokantaa nimeltä Signbank (<https://signbank.csc.fi/>), johon kootaan annotoidusta aineistosta pääasiallisesti kiinteitä viittomia glosseineen. Tietokannasta löytyy glosseilla nimetyt viittomat perusmuotoisina viitottuna, käänkösvastineet ja joitakin lisätietoja viittomista (kuvio 4). Tietokanta täydentyy koko ajan annotointiprosessin edetessä.

AIHE



Glossi:	AIHE
Glossi englanniksi:	THEME
Käännökset englanti 🇬🇧:	-
Käännökset suomi 🇫🇮:	aihe, teema, otsikko, otsake, (oppi)aine, rivi, aine(kirjoitus)
Kommentit	-
Viittomakieli:	FinSL
URL:	http://suvi.viittomat.net/wordsearch.php?a_id=898&word_search=898&offset=0&sssf=0&mpw=1
Luotu:	© 2016-02-18 09:37 🧑 Juhana Salonen
Päivitetty:	© 2017-06-23 12:04 🧑 Juhana Salonen

RIVI



Glossin relaatiot

Ei relaatioita.

Kommentit (0)

Ei kommentteja.

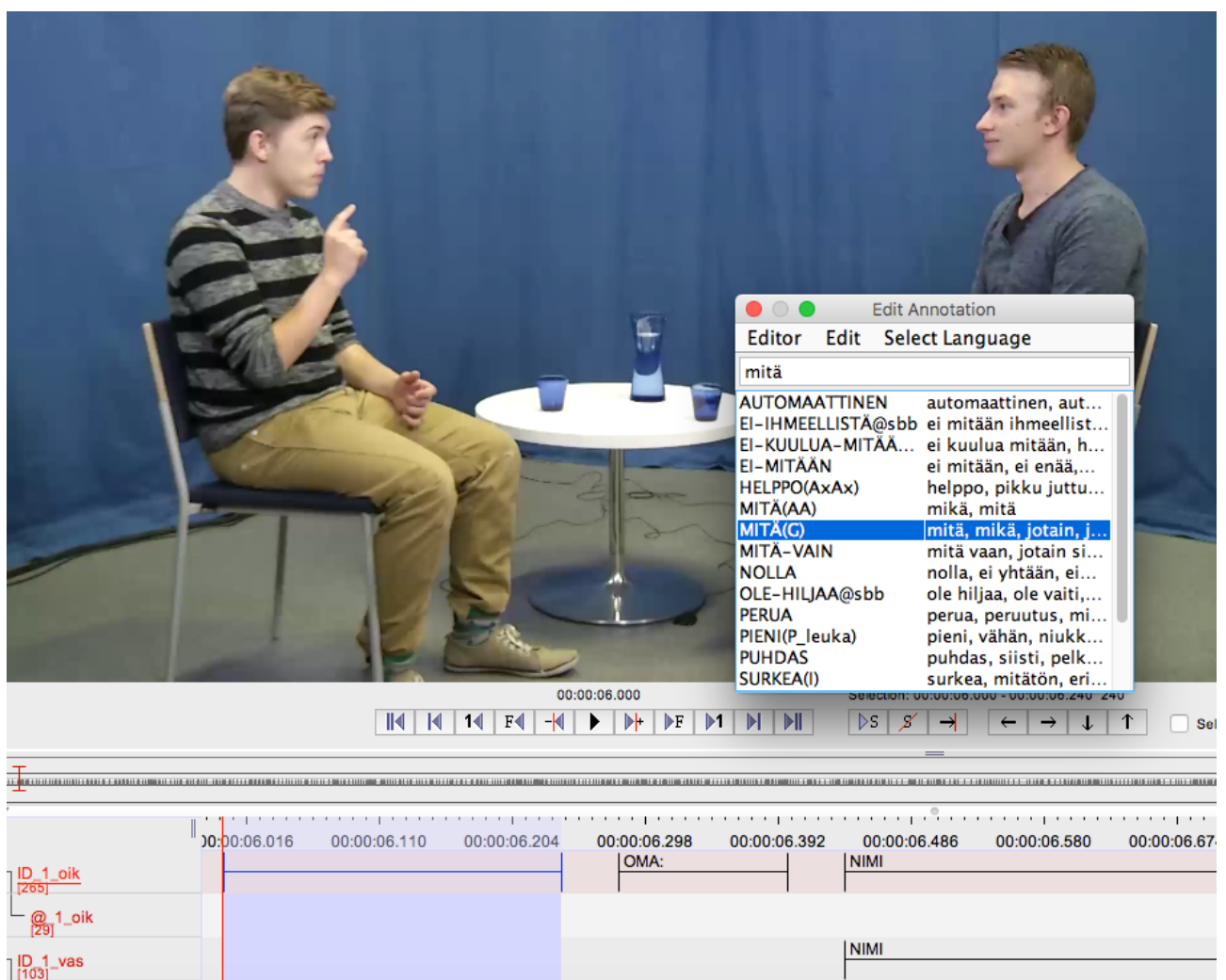
Kommentti

Kommentti

Kuvio 4. Esimerkinäkymä Signbankin glossisivusta.

Signbank on yhteydessä ELAN-ohjelmaan siten, että leksikotietokannan glossilista on verkkoyhteyden (ECV eli *external control vocabulary*) avulla saatavissa näkyviin ELAN-ohjelmassa annotoinnin aikana. Käytännössä tämä tarkoittaa, että annotoija voi joka

annotaatiolosolun nimeämisen yhteydessä tarkistaa, löytyykö glossi jo tietokannasta (kuvio 5), joko glossi tai käännöshaulla, ja valita sen sieltä. Tämä nopeuttaa ja yhtenäistää annotointia huomattavasti. Kuviossa 5 näkyy ELAN-työpöydälle Signbankista tuleva glossiluettelo, jossa näkyvät vasemmalla ID-glossit ja oikealla niiden käännösvastineet, josta annotoija valitsee viittomaan kuuluvan ID-glossin. Hakutoiminto kohdistuu myös käännösvastineisiin, joissa näkyvät viittoman merkitysvaihtoehdot. Näin annotoijan on helpompi löytää Signbankiin tallennetut viittomat ja niille annetut ID-glossit. Signbankiin tehdyt glossi- ja käännösvastinemuutokset päivittyvät automaattisesti kaikkiin linkitettyihin annotaatiolosuihin, joiden glossit ovat jatkuvassa ECV-yhteydessä tietokantaan.



Kuvio 5. Näkymä ELANin työpöydästä käytettäessä Signbankia.

Jos ID-glossia ei löydy luettelosta, tietokantaan lisätään uusi glossi sekä web-kameralla kuvattu video viittomasta. (ks. ohjeet: Keränen ym. 2018). Uudet viittomat viitotaan myöhemmin studiolla. Signbankin ID-glossisivulla voi olla useita videoita. Tällöin on kyse muoto- tai merkitysvarianteista, jotka on yhdistetty samaan ID-glossiin (ks. tarkemmin sivulta 12). Jatkossa videoiden tiedostonimissä näkyviä merkitystason glosseja voidaan hyödyntää merkitysannotoinnissa. Alla on tiivistettynä ohjeistus sille, kuinka annotointi yhteydessä Signbankiin etenee.

Annotoitaessa toimitaan seuraavasti:

1. valitaan Signbankin tarjoamasta luettelosta vastaava ID-glossi
2. tarkistetaan, että viittoman kontekstuaalinen merkitys löytyy käännösvastineista
3. tarvittaessa täydennetään käännösvastineet
4. verrataan viittoman muotoa ID-glossisivulla jo oleviin viittomiin
5. jos viittoma poikkeaa muodoltaan muista, mutta sen katsotaan kuuluvan saman ID-glossin alle, lisätään se videoituna ID-glossisivulle
6. videon tiedostonimeksi kirjoitetaan viittoman merkitys
7. jos viittomalle ei löydy ID-glossia, sitä ehdotetaan lisättäväksi Signbankiin:
 - a) viittoma viitotaan videolle
 - b) video lisätään Signbankiin
 - c) merkitään tiedoston nimeksi viittoman perusmerkitystä kuvaava sana
 - d) Signbankin glossisivulle liitetään tag: video_kuvattava_studiolla

Korpustiimi tarkistaa uuden ID-glossin ennen kuin se lopullisesti vahvistetaan ja kuvataan studiolla Signbankia varten.

4 Viittomien luokittelu ID-glossirivillä

Perusannotoinnissa tehdään alustavaa luokittelua merkitsemällä luokkatunnistin neljälle kategorialle. Kiinteillä viittomilla ei ole omaa luokkatunnistinta lukuun ottamatta numeraaliviittomia. Haun helpottamiseksi perusannotoinnissa merkitään numeraaliviittomat, kuvailevat viittomat, elemäiset viittomat ja sormiaakkoset erillisillä tunnistimilla:

1. Numeraaliviittomat (_num)
2. Kuvailevat viittomat (_kv)
3. Elemäiset viittomat (_ele)
4. Sormiaakkoset (_sa)

Seuraavaksi yksityiskohtaisemmin kuvausta näihin viittomaluokkiin kuuluvista viittomista ja niiden annotointikonventioista.

4.1 Kiinteät viittomat

Kiinteällä viittomalla tarkoitetaan leksikaalistunutta viittomaa, jonka muoto ja merkitys ovat vakiintuneet kielenkäytössä. Kiinteiden viittomien glossi määräytyy joko Signbankiin luodun ID-glossin mukaan tai sen puuttuessa sille luodaan oma ID-glossi. Kiinteät viittomat merkitään siis glossilla (käytännössä kirjoitetun kielen perusmuotoinen sana), joka kirjoitetaan isoilla kirjaimilla.

Esim. MIES, PALLO, JUODA, KÄVELLÄ¹, PUNAINEN, USEIN

Tarvittaessa glossi voi muodostua useammasta sanasta, jolloin voidaan hyväksyä myös ymmärrettävyyden kannalta välttämättömät sanataivutukset. Sanat erotetaan toisistaan väliviivalla.

Esim. MENNÄ-NUKKUMAAN, MENNÄ-KOTIIN, PIDÄTELLÄ-NAURUA, OLE-HYVÄ

Kuviossa 6 on lueteltu tarkemmin ID-glossien muodostamisen periaatteita.

¹ Emme merkitse monisuuntaisten verbaalien variantteja erikseen, pelkästään verbaalit perusmuodossa.

- a) Erimerkityksiset ja -muotoiset (väh. 1 rakenneyksikkö (parametri) eri; fonologista prosessia ei lasketa tähän mukaan) viittomat glossataan aina eri glosseilla

Esim. **RAHA** **ASIA** **LEIPÄ**

=> Jos sellainen viittoma, joka voi viitata sekä nominaaliin että verbaaliin:
Valitaan glossiksi ensisijaisesti verbi:

Esim. **KESKUSTELLA, KESKUSTELU => KESKUSTELLA**

- b) Foneettiset varianttiviittomat (on 1-2 eri rakenneyksikköä) yhdistetään 1 emoglossiksi:

Esim. **ISÄ(Ax), ISÄ(AI), ISÄ(G) => ISÄ**

OPISKELU, OPISKELLA, OPPIA(Bq), OPPIA(Ax), KURSSI(Ax) => OPISKELLA

- c) Leksikaaliset varianttiviittomat (on vähintään 2-3 eri rakenneyksikköä) erotellaan omiksi lekseemeikseen:

Esim. **AUSTRALIA(2o)** **AUSTRALIA(VI)** => käsimuototarkenteella
TEHDÄ(BB) **TEHDÄ(SS)**
PIENI(L_vasta)* **PIENI(L_alas)** => liiketarkenteella
HAUSKA(P_leuka) **HAUSKA(P_rinta)** => paikkatarkenteella

***PIENI(L_vasta)** sisältää seuraavat foneettiset variantit:
PIENI(GG)-LYHYT(GG)-PIENI(CC)-PIENI(BB)

- d) Polyseemiset ja homonymiset viittomat yhdistetään:

- Lähimerkityksisille ('polysemia'), saman muotoisille viittomille annetaan sama ID-glossi:

Esim. **PALLO, MAAILMA** => **PALLO**
SUKLAA, RUSKEA => **RUSKEA**

- Erimerkityksiset ('homonymia'), saman muotoiset viittomat merkitään samalla ID-glossilla:

Esim. **ARKI, MAASEUTU, FARKUT, RAITIS** => **ARKI**
SATAMA, RANTA, RAJA => **RAJA**

Huom! poikkeuksia esiintyy, esim. **ITSE** vs. **PARAS**

Kuvio 6. ID-glossien muodostamisen periaatteita.

4.1.1 Muita kiinteiden viittomien merkintäkonventioita

Seuraavaksi esitellään muita kiinteiden viittomien merkintätapoja. Kaikki alla mainitut merkintätavat löytyvät Signbankin tietokannasta.

- Osoitusviittomat kirjoitetaan muodossa: OS:
 => poikkeuksena minä-osoitukset kirjoitetaan muodossa: OS:minä
 => poikkeuksena B-käsimuodolla tuotettu osoitus OS(B):
 => poikkeuksena osoitus listapoijuun (ks. sivuilta 28–29)

- Omistusta ilmaiseva viittoma OMA kirjoitetaan muodossa: OMA:
 => poikkeuksena 'minun'-muodot kirjoitetaan muodossa: OMA:minun

- Refleksiivipronomini ITSE kirjoitetaan muodossa: ITSE:
 => poikkeuksena yksikön 1. persoonaan viitatessa: ITSE:minä

- Henkilöiden viittomanimet kirjoitetaan tietoturvasyistä pelkästään muodossa:
 viittomanimi

- Idiomit:
 Annotoidaan samaan tapaan kuin kiinteät viittomat. Merkitään Signbankin glossisivun käännösvastineiden perään (idiomi), jotta mahdollisia idiomeja on jatkossa helpompi löytää annotointia ja analysointia varten.

4.2 Numeraaliviittomat

Numeraaliviittomat on jaettu perus- ja inkorporoituneisiin numeraaliviittomiin. Signbankissa on valmiina perusnumeraalit (NOLLA-KAKSIKYMMENTÄ, SATA, TUHAT, MILJOONA ja MILJARDI) ja inkorporoituneet numeraalit (YKSI-KAHDEKSAN/YHDEKSÄN) inkorporoituneena johonkin kantaan, esim. KUUKAUSI. Jos kyseessä on muu perusnumero, esimerkiksi neljäkymmentäkaksi, annotoidaan se manuaalisesti kirjoittamalla ilman yhteyttä Signbankiin. Yleisesti noudatetaan sääntöä, että ensin ID-glossissa kirjoitetaan lukumäärä. Toiseksi kirjoitetaan, mikä kanta kyseessä. Kolmantena aikaan viittaava glossi ja viimeisenä _num tunnistin.

LUKUMÄÄRÄ-KANTA-MUOTOTARKENNE_num

KAKSI-VIIKKO-AIKAISEMMIN_num

Eri **numeroluokat** erotetaan toisistaan väliviivalla (nolla-yhdeksäntoista, kymmenet, sadat, tuhannet jne.). Numerot kirjoitetaan suomeksi perusmuotoisina siinä järjestyksessä kuin ne on viitottu. Numerosarjoissa otetaan kuitenkin huomioon suomen kielen ymmärrettävyyden kannalta välttämättömät taivutukset. Esim:

- TUHAT-YKSI_num
- VIISIKYMMENTÄ-KOLME_num
- TUHAT-YHDEKSÄNSATAA-KAHDEKSANKYMMENTÄ-KAKSI_num
- YHDEKSÄN-SATAA_num

Kun kanta on johdettu jostain viittomasta, **inkorporoituneet numeraalit** annotoidaan seuraavasti:

- KUUSI-VUOSI_num
- NELJÄ-VIIKKO_num
- KOLME-KUUKAUSI_num
- KAHDEKSAN-KERROS_num

Tarvittaessa lisätään **viittoman muotoa tarkentava glossi**. Esim:

- VIISI-VIIKKO-MYÖHEMMIN_num
- VIISI-VIIKKO-AIKAISEMMIN_num
- KUUSI-VUOSI-MYÖHEMMIN_num
- KUUSI-VUOSI-AIKAISEMMIN_num

Järjestysnumerot kirjoitetaan sellaisenaan tarvittavien tarkenteiden kanssa. Esim:

- YHDEKSÄS_num
- TOINEN(L_alas)_num
- TOINEN(L_sivulle)_num

Huom! a)

Niissä tapauksissa, joissa ei voida tunnistaa viittomaa, josta kanta on otettu, merkitään glossiksi ko. NUMERAALI-MERKITYS_num
(Esim. KAHDEKSAN-KERROS_num, NELJÄ-KELLO_num, KAKSI-LUOKKA_num).

Huom! b)

Jos käytetään VIIKKO -viittomaa, jossa on käsimuotona Lc, se ei ole numeraaliviittoma, vaan kiinteä viittoma, jolloin se annotoidaan esimerkiksi muodossa:
VIIKKO(L_taakse).

4.3 Kuvailevat viittomat

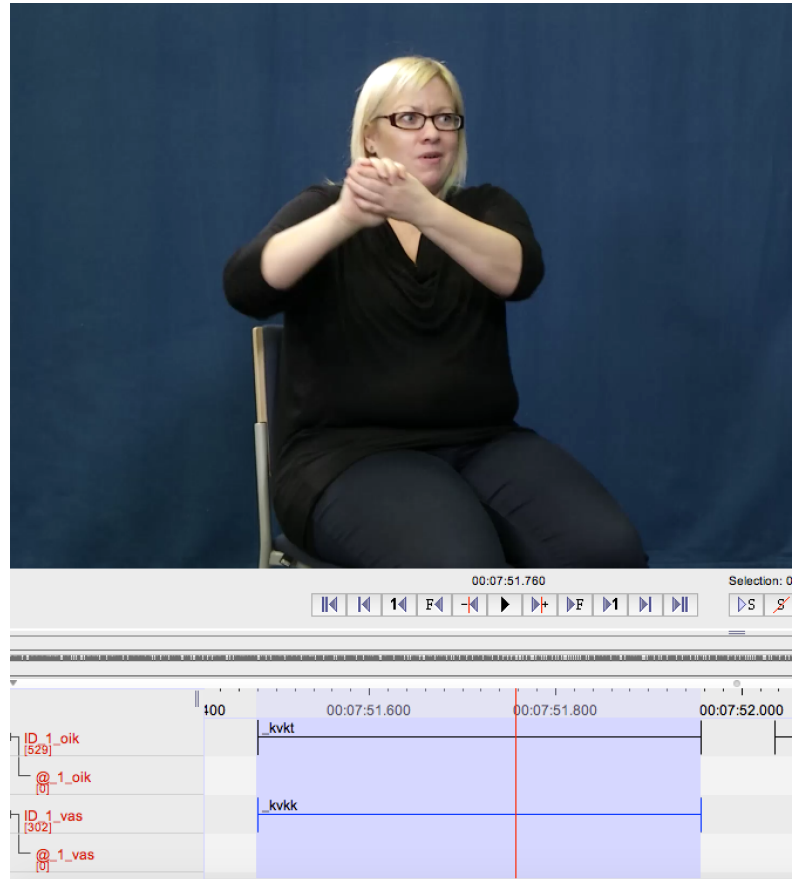
Kuvailevat viittomat on jaettu kuuteen alaluokkaan (joista kolme ensimmäistä Takkiselta (2008)), jotka on lueteltu taulukossa 1. Kuvaileva viittoma merkitään perusannotoinnissa tunnistimella _kv ja Signbankista valitaan kummallekin kädelle sitä kuvaava alaluokan tunnistin (ks. kuvio 7). Kuvailevien viittomien tarkempi kuvaus tehdään myöhemmässä vaiheessa Merkitys- ja Käännös-riveille.

Käännös:

Tarttua lumiukon käteen.

ID_1_oik: _kvkt

ID_1_vas: _kvkk



Kuvio 7. Kuvailevien viittomien annotointiesimerkki.

Taulukko 1: Kuvailevien viittomien luokittelu

<u>KUVAILEVAN VIITTOMAN TYYPPI</u>	Kuvailevan viittoman tunnistin
1) Kokonainen kohde	_kvkk
2) Käsittely	_kvkt
3) Muoto ja koko	_kvmk
4) Aika ja paikka	_kvap
5) Abstraktinen kiintopiste	_kvak
6) Epäselvät tai luokittelemattomat tapaukset	_kvx

ESIMERKKEJÄ:**1) Kokonainen kohde:**

Käännös: Auto kaahaa eteenpäin.

ID_1_oik: _kvkk

Käännös: Auto törmää ihmiseen.

ID_1_oik: _kvkk

ID_1_vas: _kvkk

2) Käsittely:

Käännös: Lasipurkki juuttuu päähän.

ID_1_oik: _kvkt

ID_1_vas: _kvkt

Käännös: Nostaa omenan maasta.

ID_1_oik: _kvkt

3) Muoto ja koko:

Käännös: pitkä puu

ID_1_oik: _kvmk

ID_1_vas: _kvmk

Käännös: Ruokaliina rinnalla.

ID_1_oik: _kvmk

ID_1_vas: _kvmk

4) Aika ja paikka:

Käännös: Aikaa kuluu.

ID_1_oik: _kvap

ID_1_vas: _kvap

Käännös: Etäisyys näiden välillä.

ID_1_oik: _kvap

ID_1_vas: _kvap

5) Abstraktinen kiintopiste

Käännös: poika juoksee alas (ks. kuvio 8).

ID_1_oik: OS:

ID_1_vas: _kvak

(Tässä vasemmalla kädellä tuotettu käsimuoto ei viittaa varsinaisesti seinään. Se on ikään kuin abstraktina kiintopisteenä, joka hahmottaa oikean käden viittaaman olennon liikkumista.)

Käännös: Jäädä pois.

ID_1_oik: _kvkk

ID_1_vas: _kvak

(Tässä samaan tapaan vasemmalla kädellä tapahtuva abstraktinen kiintopiste toimii ikään kuin taustana tai kiintopisteenä. Se ei viittaa mihinkään konkreettiseen entiteettiin, vaan lähtöpisteeseen.)



Kuvio 8. Esimerkki vasemmalla kädellä tuotetusta abstraktisesta kiintopisteestä.

6) Epäselvät tapaukset

Käännös: Hyökätä jonkin kimppuun...

ID_1_oik: _kvx

ID_1_vas: _kvx

Käännös: Aistia informaatiota silmien kautta...

ID_1_oik: _kvx

ID_1_vas: _kvx

4.4 Eleet

Eleen ja viittoman erottelu ei aina ole helppoa. Eleet voidaan jaotella jatkumolle alkaen täysin tahattomista elehtimisistä ja päättyen hyvinkin vakiintuneisiin embleemeihin. Perusannotoinnissa sellaiset esiintymät, joita ei voida luokitella kuuluviksi kiinteisiin viittomiin, sormiaakkosiin, eikä kuvaileviin viittomiin, merkitään _ele -tunnistimella. Eleiden koodaaminen aineistosta on tärkeää, jolloin nekin ovat hakutoimintojen avulla saatavissa tarkempaa analysointia varten. Tässä vaiheessa tehdään karkeaa jakoa vakiintuneisiin eleisiin ja ei-vakiintuneisiin eleisiin. Aineistosta löytyneille, vakiintuneessa käytössä oleville eleille on luotu omat ID-tunnukset, ja ne on lisätty Signbankiin.

Vakiintuneessa käytössä oleville eleille valitaan Signbankista siihen viittaava glossi ja tarvittaessa lisätään uusi ele tietokantaan. Muiden ei-vakiintuneiden eleiden (elehtiminen, pantomiimi) kohdalla valitaan glossiksi pelkkä _ele, joka on lisätty Signbankiin. Vakiintuneet eleet merkitään samaan tapaan isoin kirjaimin kuin kiinteät viittomat sekä lisätään tunnistin _ele glossin perään. Signbankin tietokannassa on tällä hetkellä esimerkiksi:

HEI-KUULE_ele
 KÄMMEN-YLÖS_ele
 KÄMMEN-ALAS_ele
 KÄSI-YLÖS_ele
 TUUMIA_ele

4.5 Sormiaakkoset

Sormiaakkoset annotoidaan aina manuaalisesti, pienillä kirjaimilla, tavuviivalla erotettuina ilman yhteyttä Signbankiin. Sormitusjakson loppuun merkitään _sa (esim. p-e-k-k-a_sa). Tiedot sormitetuista jaksoista merkitään ainoastaan annotointiriville. Sen sijaan sormiaakkosista muodostuneet viittomat (esim. viittoma OK) lisätään leksikotietokantaan.

Sormiaakkoset annotoidaan täsmälleen samassa muodossa kuin mitä omin silmin nähdään. Kesken jäänyt sormitus annotoidaan siis sellaisenaan ja tytärriville merkitään @kes. Käännösriville voidaan kirjoittaa tarkemmin sormituksella tarkoitettu merkitys.

- t-o-m-i_sa (perään liitetään _sa)
- k-a_sa (kesken jäänyt sormitus nimestä *Katja*) + tytärriville:
@kes
- zzz_sa (epäselvä sormitus)
- w-e-d-e-zzz_sa (alkuosa *wede*, minkä jälkeen epäselvää)

Huom! Jos sormituksessa käytetään vanhoja sormiaakkosia, merkitään siitä ilmoitus Huomioita -riville.

5 Viittoman lisätiedot

Edellä esitettiin tunnistimia (esim. _ele), joita käytetään viittomien luokittelussa. Sellaisten viittomien erotteluun, joiden suomenkieliset glossit ovat identtiset (ks. 5.1), ja erilaisten kieliopillisten piirteiden (ks. 5.2) merkitsemiseen tarvitaan myös koodeja.

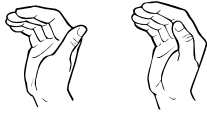
5.1 Rakenneyksikkötietojen merkitseminen ID-glossiin

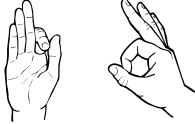
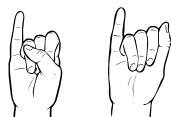
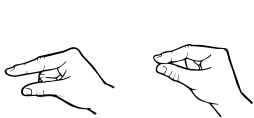
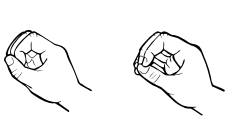
Leksikaalisten varianttiviittomien eli synonyymien kohdalla tarvitaan koodeja, joiden avulla merkitään varianttien muotoeroja. Erotteluun käytetään koodeja neljälle rakenneyksikölle: käsimuoto, liike, orientaatio ja paikka. Jos vaihtelua on useissa rakenneyksiköissä, valitaan merkittäviksi ensisijaisesti käsimuodon koodit. Jos viittomien käsimuodoissa ei ole eroa, valitaan jokin toinen, selvimmin havaittava rakennepiirre merkittäväksi omalla koodillaan. Lisäksi voidaan merkitä tietoa siitä, onko kyseessä yksi- vai kaksikäsinen viittoma.

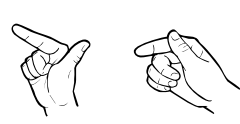
Käsimuodoille käytetyt koodit löytyvät olevasta taulukosta. Käsimuotojen piirroskuvat on saatu Kuurojen Liitto ry:n sanakirjatyöstä.

Taulukko 2. Käsimuotojen koodit ja niiden piirroskuvat.

Koodi (Tyyppi)	Koodia edustavat käsimuodot
A	
Al	
Ax	

B	
Bc	
Bq	
C	

D	
E	
F	
G	
Gi	
H	
I	
K	
L	
Lc	
Lo	
Lq	
M	
O	

R	
S	
T	
V	
Vc	
VI	
W	
X	
Y	
2	
2o	
4	
5	
5c	



Esimerkkejä ID-glosseista, joihin on lisätty tietoa käsimuodoista:

AUSTRALIA(2o)	vs.	AUSTRALIA(VI)
LIPPU(AxB)	vs.	LIPPU(B)
PÄÄ(B)	vs.	PÄÄ(G)
KAHVI(SS)	vs.	KAHVI(F)

Liikkeiden erottelussa käytetään glossin perään lisättyä koodia L_, jonka jälkeen tulee tarkempi kuvaus liikkeen suunnan tai tavan mukaan.

Esim.

JÄRVI(L_ympyrä)	vs.	JÄRVI(L_loittoneva)
AJAA-AUTO(L_eteen)	vs.	AJAA-AUTO(L_vuoro)
TIETÄÄ(L_toisto)	vs.	TIETÄÄ(L_loiva)
TUTKIA(L_alas)	vs.	TUTKIA(L_ympyrä)

Orientaation vaihtelussa huomioidaan kämmenten suunta. Käytämme koodia O_ glossin perässä, jonka jälkeen määritellään kämmenen orientaation suunta.

Esim.

EI(O_ylös)	vs.	EI(O_eteen)
KATSOA(O_eteen)	vs.	KATSOA(O_takse)
VAPAA(O_alas)	vs.	VAPAA(O_ylös)

Paikan tarkenteena käytämme koodia P_ glossin perässä, jonka jälkeen kuvaillaan tarkemmin viittoman paikkaa.

Esim.

HAUSKA(P_leuka)	vs.	HAUSKA(P_rinta)
KIRJOITTAA(P_kämmen)	vs.	KIRJOITTAA(P_neutraali)
KUURO(P_korva)	vs.	KUURO(P_suu)

5.2 Kieliopillisten piirteiden merkitseminen tytäriville

CFINSL-projektin annotointi on ID-glosseihin pohjautuvaa perusannotointia. Joitakin kieliopillisia piirteitä on kuitenkin myös tässä vaiheessa huomioitu. Kun viittomasta halutaan kirjata lisätietoa jostakin kieliopillisesta piirteestä, käytetään @-koodia.

@-koodin idea on peräisin ruotsalaisen viittomakielen korpustutkimuksesta (Wallin & Mesch 2018). Se on osoittautunut tehokkaaksi hakutyövälineeksi tietyn luokan etsimiselle, minkä lisäksi @ on helppo erottaa muista kirjaimista ja merkeistä. @-koodi kirjoitetaan tarkentavine tietoineen tytäriville, joka on emorivin (ts. ID-rivin) alla ja jatkuvassa yhteydessä siihen. Tytäriville tehtävät merkinnät toteutetaan manuaalisesti, eikä niitä siirretä Signbankiin. Seuraavaksi esitellään ne kieliopilliset piirteet, jotka on otettu käyttöön perusannotoinnissa.

5.2.1 Kielto

a) *Verbaaliviittoma + kieltojohdin tai kielteisen merkityksen sisältävä verbaaliviittoma*

Jos kyseessä on kieltoviittoma, joka on muodostettu kieltojohtimella tai viittoma on merkitykseltään kielteinen, kirjoitetaan glossi seuraavasti: VIITTOMAN PERUSMERKITYS-EI. Tällaiseen ratkaisuun on päädytty hakuteknisistä syistä, koska nyt hakutulos antaa

verbaaliviittoman sekä myönteisen että kielteisen muodon allekkain, mikäli sellaisia on. Lisäksi kirjataan kategorinen merkintä @neg tytärriiville.

Kieltojohdos:

ID_1_oik TIETÄÄ-EI
 @_1_oik: @neg
 ID_1_vas TIETÄÄ-EI
 @_1_vas: @neg

Kieltoviittoma:

ID_1_oik TYKÄTÄ-EI
 @_1_oik: @neg

Muita esimerkkejä:

Kieltojohdoksia	Kieltoviittomia
OLLA-EI	USKALTAA-EI(P_kämmen)
HALUTA-EI	VOIDA-EI

b) Kielto päänpudistuksella

Jos kielto ilmaistaan pelkällä päänpudistuksella, eikä manuaalista kieltoa ole: kirjoitetaan verbaalille valittu glossi ID-riville ja tytärriiville @pp (päänpudistus) ja sen perään @neg.

=> ID-glossirivi SYÖDÄ TULLA
 => Tytärriivi @pp@neg @pp@neg

c) *Kieltopartikkeli*

Jos kyseessä on 'partikkelin' tapainen ilmaisu (ei mitään, ei vielä, ei koskaan...), glossi kirjoitetaan suomen kielen mukaisessa sanajärjestyksessä väliviivalla erottaen (EI-MITÄÄN). Näiden viittomien kohdalla tällainen käytäntö on mielekäs myös hakutoimintojen näkökulmasta, sillä sanajärjestyksen vaihtaminen ei anna mitään lisäarvoa, toisin kuin verbaalin sisältävissä kieltoviittomissa, joilla on myös myönteinen muoto.

partikkeliviittoma =>	EI-MITÄÄN	EI-KOSKAAN
	@neg	@neg

Signbankiin lisätään kieltojohdos-, kieltoviittomat ja kielteiset partikkelit.

5.2.2 Monikko ja toisto

ID-glossirivien tytär-riveille merkitään @toisto silloin kun viittomaa on toistettu monikon tai toisteisuuden ilmaisemiseksi nominaali-, verbaaliviittomissa ja luonnehtivissa statiivisissa viittomissa. Rissanen (1985) mukaan nominaaliviittomia tai niiden liikejuurta voidaan toistaa monikollisuuden ilmaisemiseksi. Verbaalien ja luonnehtivien statiivisten viittomien kohdalla voidaan toistolla ilmaista esimerkiksi toiminnan kestoa tai useutta. (Rissanen emt.) Jos viittoman perusmuoto sisältää toistoliikeyä, sitä ei merkitä toistoksi (esim. ASIA). Toisto merkitään vain siinä tapauksessa, jos liikettä on **lisätty**. Glossia ei kirjoiteta monikolliseen muotoon. Esimerkiksi:

					
ID-rivi:	TALO	TAVATA		VIHAINEN	
Tytärrivi:	@toisto	@toisto		@toisto	
Käännös:	'taloja'	'tavata usein'/ 'tavata harvoin'		'olla pitkään vihainen'	

Kuvio 9. Viittomat TALO, TAVATA ja VIHAINEN, joissa esiintyy toistoa ja niiden annotointi. Kuvat on poimittu lähteestä Rissanen (1985: 107, 113 ja 116).

Poikkeuksena ovat suomen kielen monikkomuotoiset sanat, jotka kirjoitetaan sellaisenaan, vaikka viittoma ei ole monikollinen. Tällöin glossiksi kirjoitetaan monikkomuotoinen sana ilman @toisto -merkintää tytärriville. Esimerkiksi:

HOUSUT SAKSET KULTTUURIPÄIVÄT

5.2.3 Listapoijut

Poijuista annoimme pelkästään listapoijut (merkitään muodossa @poijul). Listapoiju tuotetaan useimmiten ei-dominoivalla kädellä ja dominoivalla kädellä viitataan siihen yleensä osoittamalla (esim. kuvio 10). Signbankiin on lisätty seuraavat glossit: OS:peukalo, OS:etusormi, OS:keskisormi, OS:nimetön ja OS:pikkusormi, joista valitaan se vaihtoehto, mihin kulloinkin osoitetaan. Numeroiden sijasta käytämme sormikohtaisia termejä, koska numeroiden järjestys voi vaihdella (esim. peukalo voi tarkoittaa joko yhtä tai viittä).

Kulloinenkin merkitys on luettavissa Merkitys- ja Käännösriviltä. Poijua tuottavan käden riville kirjoitetaan ojennettujen sormien lukumäärä ja kirjataan perään _num, sekä sen tyttärriville merkitään @poijul.

Dominoiva käsi (oikea):

- mihin sormeen osoitetaan

Ei-dominoiva käsi (vasen):

- montako sormea on listassa
- tyttärriville @poijul

Viereisen kuvan esimerkissä:

ID_1_oik OS:nimetön

ID_1_vas KOLME_num

@_1_vas @poijul

.



Kuvio 10. Esimerkki listapojun annotoinnista.

5.2.4 Yhdysviittomat

Yhdysviittoman määrittely ei ole yksiselitteistä, sillä tutkimukset antavat erilaisia määritelmiä. Jokainen potentiaalinen yhdysviittoma merkitään kuitenkin Huomioita -riville (**ei siis tytärriville**) koodilla @y, jolloin ne on haettavissa jatkotarkastelua varten. Perusannotointivaiheessa ei myöskään ole tarpeen viedä kaikkia mahdollisia viittomayhdistelmiä ominaan Signbankiin. Joidenkin selvien yhdysviittomien kohdalla on päädytty annotoimaan seuraavasti:

Yhdysviittoma **annotoidaan yhteen soluun kun:**

1. Yhdysviittoman osat ovat sulautuneet yhteen ja viittoma näyttää lähes yhdeltä viittomalta sekä muodostaa uuden käsitteen. Tällöin kirjoitetaan glossiksi vastaava suomenkielinen käännösvastine. Jos ID-glossiksi tarvitaan useampia sanoja, ne erotetaan väliviivalla. Esimerkiksi:

VIITTOMAKIELI	AURINKO
TIETOKONE	KUUROJENYHDISTYS
MENNÄ-KOTIIN	MENNÄ-ULOS

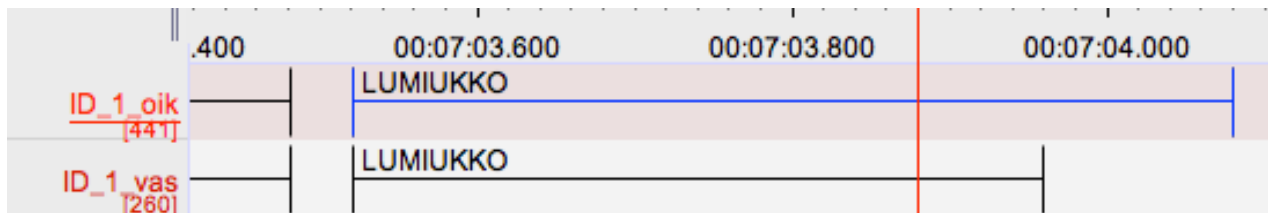
2. Ilmauksesta on erotettavissa kaksi viittomaa, mutta kyseessä on yksi käsite. Kriteereitä ovat esimerkiksi:
 - a. yhtenäisyys: viittomien välissä ei ole välttämättä silmällä erotettavaa taukoa
 - b. ei-dominoiva käsi on mukana viittoman tuottamisessa molemmissa osissa, vaikka toinen yhdysviittoman osa olisi yksikäsinen viittoma (kuvio 12)
 - c. yhdysviittoman osat tuotetaan lähes poikkeuksetta samassa järjestyksessä
 - d. aineistossa yhdistelmän frekvenssi on suuri (LUMIUKKO, YLIOPISTO)

- e. yhdistelmän tiedetään olevan yleinen kuurojen kulttuurissa tai erilaisissa viitotuissa kielenkäyttötilanteissa (KUUROJEN-LIITTO, KUUROJENKOULU, KULTTUURIPÄIVÄT).
- f. Erisnimet, kuten paikat, henkilöt ja muut tahot (HAAPAMÄKI, HAUKKARANTA, KAUPPAKATU)
- g. Idiomaattiset ilmaukset (TAJUTA(55), YKSINKERTAINEN, ITSEPÄINEN), joita kuitenkin käydään läpi yksittäin.

Tämän tyyppiset yhdysviittomat lisätään tilannekohtaisen tarkastuksen jälkeen omina kokonaisuuksinaan Signbankiin.

Huomioitava:

Kaksikättiset yhdysviittomat, joille on luotu oma ID-glossi, annotoidaan siten, että molempien käsien kohdalla on aina sama ID-glossi. Näin niissäkin tapauksissa, kun kaksikättisesti on tuotettu vain yhdysviittoman ensimmäinen tai jälkimmäinen osa (esim. LUMIUKKO, ks. kuvio 12). Annotaatiosolujen pituus määrittyy kunkin käden aktiivisuuden mukaan (ks. kuvio 11 tarkemmin).



Kuvio 11. Esimerkki yhdysviittoman annotoinnista.



Kuvio 12. LUMIUKKO viitottuna.

Muutoin annotoidaan viittomat **omiin soluihinsa** (esim. KANOOTTI LEIRI). Myös viittomaparit, joista toinen viittaa tekijään (esim. AJAA HENKILÖ 'kuljettaja') tai kyseessä on vertailumuoto (esim. KAUNIS ENSIMMÄINEN_num 'kaunein'), kirjoitetaan eri annotaatiolosuihin. Huomioita -riville laitetaan merkintä @y näissäkin tapauksissa.

5.2.5 Sormiaakkosia sisältävät viittomat

Sormiaakkosia sisältävät viittomat muodostuvat kolmesta eri ryhmästä (Jantunen 2003): leksikaalistuneista sormituksista, alkukirjainviittomista ja sormiaakkosia sisältävistä yhdysviittomista.

Leksikaalistuneet sormitukset sisältävät vähintään kaksi sormiaakkosta, jotka on yhdistetty liikkeellä (esim. TV, TIISTAI, OK, EU). **Alkukirjainviittomissa** viittoma tuotetaan sormiaakkoskäsimuodolla. Sormiaakkoskäsimuoto viittaa kirjoitetun sanan ensimmäiseen kirjaimeen (esim. IDIOMI, TUHAT, PERJANTAI, TEE). **Sormiaakkosia sisältävät yhdysviittomat** koostuvat viittomista, joiden alkuosassa esiintyy sormiaakkoskäsimuoto. (Jantunen emt.) Ne annotoidaan omiin soluihinsa ja Huomiota -riville merkitään @y.

Esim. E-PILLERI:	e_sa PILLERI
ANTIBIOOTTI:	a_sa OHJATA
VIRUS:	v_sa VAIKUTTAA

Tytärriville merkitään koodi @sv sormiaakkosia sisältävien viittomien kohdalle, jolloin ne ovat haettavissa tarkempaa analyysiä varten.

5.2.6 Lainaviittomat

Jos viittoma on tunnistettavissa lainaviittomaksi toisesta viittomakielestä, eikä ole vakiintunut suomalaisen viittomakielen leksikkoon, kirjoitetaan glossi lähtökielen mukaisella kielellä ja tytärriville merkitään @lv. Jos annotoija on epävarma mahdollisen lainaviittoman alkuperästä, hän voi kirjoittaa kommentin siitä Huomioita -riville. Signbankin glossisivulle merkitään lisätietojen kohtaan koodi @lv ja tiedot viittoman alkuperästä ja sen

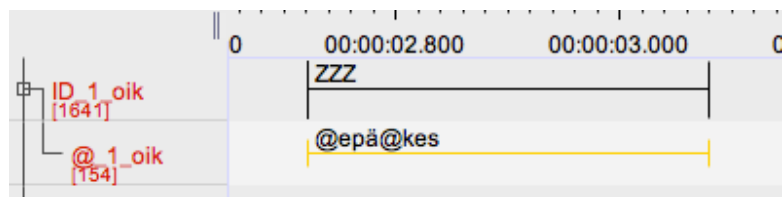
merkityksestä.

Esimerkiksi POWER amerikkalaisesta viittomakielestä; FRITID ruotsalaisesta viittomakielestä tai FAVELA brasilialaisesta viittomakielestä.

5.2.7 Epäselvät viittomat

Annotoinnissa tulee vastaan ilmauksia, joiden merkityksestä tai muodosta ei ole täyttä varmuutta. Lisäksi ilmauksia, joissa viittomaa haetaan tai se keskeytyy. Näissä tapauksissa valitaan ID-glossiriville glossiksi ZZZ, joka on valmiina Signbankissa. Glossin alle tytärriville valitaan yksi tai useampi vaihtoehto seuraavista:

@epä = viittoman muoto tai merkitys on epäselvä
 @hv = hakea viittomaa
 @kes = keskeytynyt viittoma



Kuvio 13. Esimerkki epäselvästä ja keskeytyneen viittoman annotoinnista.

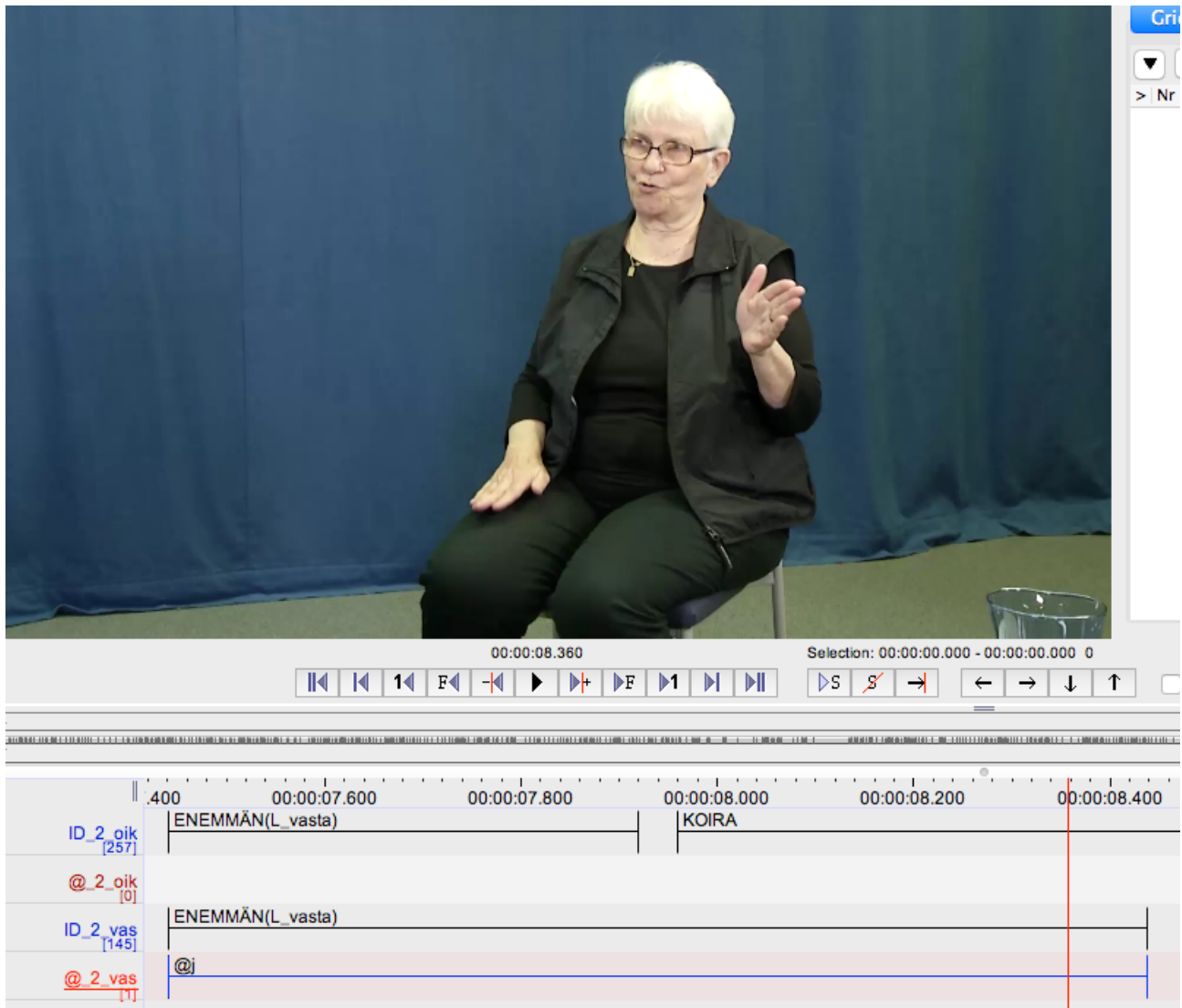
Huom! Koodi @kes käytetään myös silloin kun viittoma on tunnistettu, mutta se jää kesken. Tarvittaessa voi kirjoittaa lisäkommentteja Huomioita -riville.

5.2.8 Käden jähmettyminen

Jos jompikumpi käsistä jähmettyy paikoilleen merkittävästi pidemmäksi aikaa (**enemmän kuin 5 videoruudun ajan**) viittoman tuottamisen jälkeen, merkitään tytärriville koodi @j. Käden jähmettyminen voi viitata moniin seikkoihin, kuten esimerkiksi merkityksellisellä tasolla fragmenttipoijuun tai ei-merkityksellisellä tasolla käden 'unohtumiseen'.

Ainoastaan pidempikestoiset jähmettymiset merkitään oman intuition pohjalta jatkokäsittelyä varten. Lyhyet alle 5 videoruudun jähmettymiset jätetään huomioimatta.

Esim. vasemmalla kädellä tuotetun ISO -viittoman (ID-glossina ENEMMÄN(L_vasta)) jähmettyminen (kuvio 14):



Kuvio 14. Käden jähmettyminen.

5.2.9 Pään nyökkäys ja pudistus

Jos viittoja ei käytä lainkaan manuaalisia artikulaattoreita, vaan nyökkää tai pudistaa päätään ilmaistakseen myönteisen tai kielteisen merkityksen, merkitään Huomioita -riville koodi @pn, jos on kyse pään nyökkäyksestä. Vastaavasti @pp kun on kyse päänpudistuksesta. Huom! Keskustelun ylläpitämiseen tarkoitettuja nyökkäyksiä ja päänpudistuksia ei oteta huomioon. Emoriveille (sekä oikea että vasen käsi) merkitään ZZZ.

6 Käännöksen periaatteita

Korpusprojektissa on tavoitteena saada videoaineistoihin perusglossauksen lisäksi käännökset. Suomalaista viittomakieltä sisältävät aineistot käännetään suomen kielelle ja tavoitteena on käännös, joka kertoo viitotun asiasisällön noudatellen lähtökielen tapaa ilmaista asiat sekä huomioi myös ei-manuaalisesti ja muilla kuin leksikaalisilla tavoilla ilmaistuja asioita. Tällä tavalla saadaan myös kirjoitetussa muodossa kokonaisvaltaisempi kuva aineistossa olevista viitotuista teksteistä.

Käännöksen lisäksi aineistolle tehty ID-glossaus on tehty ainoastaan tekstien manuaalisesti tuotetuille osille ja ei-leksikaalinen aines, kuten kuvailevat viittomat ja eleet on identifioitu ja luokiteltu hyvin yleisellä tasolla. ID-glossien tarkoitus on ollut nimetä samanmuotoiset viittomat hakutoimintoja varten, eivätkä näin ollen ole käännöksiä. Luettelon mahdollisista käännösvastineista saa, kun siirtää hiiren ID-glossin päälle. Tällöin tulee näkyviin ne käännösvastineet, jotka viittomalle on siihen mennessä kirjattu.

Alla oleva ohjeistus auttaa kääntäjiä toimimaan yhdenmukaisemmin, sekä käännösten lukijaa tulkitsemaan suomen kielen kirjoituskonventioista poikkeavia merkintöjä. Käännökset antavat mahdollisuuden myös laajempiin hakutoimintoihin.

Ensimmäinen vaihe on **erotella viitotusta tekstivirrasta mielekkäitä ilmauskokonaisuuksia**. Tavoitteena ei ole erotella tarkasti *lauseita*, sillä lause ja sen määrittely ovat varsinaisen tutkimuksen tehtävä. Ilmauskokonaisuuksien erottelu on siis kunkin kääntäjän päätettävissä. Yksi ilmaus voi siis sisältää esim. pitkiäkin verbaaliketjuja. Huomioitava on viittomakielinen tapa esim. toistaa verbaalia, eli ilmaus saattaa sisältää kaksi kertaa saman verbaalin, mutta se ei välttämättä näy käännöksessä.

Vapaassa keskustelussa tai kerronnassa on luonnollista, että viittoma, lause tai ilmaus keskeytyy. Tällaisissa tapauksissa se on merkitty käännöksiin näkyviin kolmena pisteenä (...).

Pääsääntöisesti pyritään tekemään tekstistä adekvaatteja käännösvastineita ottaen huomioon kokonaisvaltainen ilmaus, mukaan lukien ei-manuaalisin ja tilankäyttöisin keinoin

ilmaistut asiat. Kuitenkin täytyy pitää mielessä, että käännöksiä ei ole tarkoitus käyttää sellaisinaan, vaan yhdessä videoaineiston kanssa. Sellaisissa tapauksissa, joissa viittoja näyttää jotakin (esim. kokoa tai muotoa), voidaan kääntää esimerkiksi: näin, tällainen, tällä tavalla.

Tämän vaiheen käännöksissä pyritään olemaan lähtötekstiuskollisia. Käännöksissä saa (ja pitääkin) näkyä viittomakielinen tapa ilmaista asioita ja ilmiöitä. Käännöstekstin tulee olla luettavaa, jonka vuoksi niihin on tehty lisämerkintöjä. Yleisesti voidaan todeta, että sulkein merkityt sanat eivät ole ilmaistu erillisin manuaalisesti tuotetuin viittomin ja yhdysviivalla yhdistetyt sanat viittaavat 1) yhteen viittomaan, 2) yhtäjaksoisesti esitettyyn tapahtumaketjuun, tai 3) samanaikaisesti ilmaistuihin viittomiin (kummallakin kädellä tuotettu oma merkitys).

Merkinnät tuovat esille suomalaisen viittomakielen ja suomen kielen erilaisia tapoja ilmaista asioita ja ilmiöitä. Tarkoitus ei kuitenkaan ole mennä niin pitkälle, että esimerkiksi suomen kielen taivutuksia tai päätteitä eroteltaisiin yksittäisistä sanoista.

Alla muutamia yksittäisiä ohjeita:

Aikamuodon valinta: Jos viitotussa tekstissä ei ole viittomaa, jolla viitataan esim. menneisiin tapahtumiin (SILLOIN-ENNEN, JO, VALMIS, EILEN jne.), tai kontekstista ei ole selvästi pääteltävissä, että viitataan menneisiin tapahtumiin, käännöksessä käytetään preesensiä. Pitemmissä teksteissä voi esiintyä eri aikamuotoja.

Kopula: Suomen kielessä lauseet useimmiten edellyttävät verbin mukanaoloa, kuten esim. olla-verbiä. Suomalaisessa viittomakielessä tilanne on toinen. Suomen kielen kannalta olennaiset verbit lisätään ja ne kirjoitetaan sulkuihin.

OS HOTELLI => Tuossa (on) hotelli.

Tekijän merkitseminen: Suomen kielessä aktiivilauseissa ilmaistaan subjekti omalla sanallaan tai päätteellä. Suomalaisessa viittomakielessä tekijä ilmaistaan myös monin muun tavoin. Käännöksen edellyttämä subjekti merkitään sulkuihin niissä tapauksissa, kun

alkuperäisessä tekstissä ei ole sille viittomin ilmaistua vastinetta. Subjektiksi valitaan tekstin yhteydessä oleva tekijä, joka on mainittu aiemmin tai joka ilmenee asiayhteydestä.

Esim. ...SYÖDÄ LEIPÄÄ => ...(nainen) syö leipää / (hän) syö leipää.

Osoitusten kohdalla käännösvastineeseen voidaan lisätä tarkennus pelkän pronominin lisäksi.

Esim. OSOITUS-3 SYÖDÄ => Hän (poika) syö.

Joissakin tapauksissa myös passiivilauseet ovat hyviä käännösratkaisuja.

Konstruoitu toiminta (CA): Erityisesti narratiiviteksteissä esiintyy usein CA:ta, eli viittoja näyttää tai esittää omaa tai jonkun toisen toimintaa, tekemistä tai reaktiota. Yleensä ennen CA-osuuden alkua viittoja ilmaisee viittomin tai osoittamalla kenestä on kyse. Tämä näkyy myös käännöksissä. Näille, usein hyvin informatiivisille CA-ilmauksille ei löydy yhtä käännösvastinetta, vaan niihin voidaan tarvita useita suomen kielen vastineita. Nämä erotellaan toisistaan yhdysviivalla (-), jolloin ilmaistaan, että kyseessä on yksi viittoma, tai yhtenäinen viittomajakso. Esim:

KÄVELLÄ (+raskas liike & ilme) => kävelee-raskaasti-eteenpäin, tai kun viittoja esittää montaa tapahtumaa samanaikaisesti: pitelee-lehteä-(ja)-lukee-sitä-ääneen.

Viittoja on aiemmin ilmaissut kiinteällä viittomalla, että kyseessä on poika ja tarinan edetessä CA:n keinoin viitto:

KULHO kauhoa-jollakin-suuhunsa-litteältä-alustalta => (Poika/hän) kauhoo-ruokaa-suuhunsa-kulhosta-siitä-pidellen.

Referointi: Jos viittoja referoi itsensä, esittämänsä hahmon tai jonkun muun sanomisia tai ajatuksia, niin niiden käännökset merkitään lainausmerkkien (" ") sisään. Esim.

Ei (toistoa x3) => (nainen): "Ei, ei, ei!"

OS:minä HALUTA-EI => (Työkaveri): "En-minä-viitsi."

Monisuuntaiset viittomat: Joihinkin viittomiin voidaan koodata tietoa viittoman liikkeen suuntaamisella (tekijä, kohde, suunta). Tällöin tämä lisämerkitys näkyy käännöksessä siten, että se on myös eroteltu yhdysviivalla, jolloin lukijalle tulee selväksi, että kyseessä on yksi viittoma. Esim.

KÄVELLÄ => kävelee-pois-paikalta.

ANTAA => antaa-hänelle.

NORSU KATSOA => Norsu katsoo-minuun.

Konjunktiot: Viittomakielessä on useita tapoja ilmaista esim., että kyseessä ovat pää- ja sivulause, tai lauserinnastukset. Suomen kielen tapa on käyttää useimmiten konjunktioita. Käännöstekstissä konjunktiot kirjoitetaan sulkuihin, jos niitä ei ole erillisellä viittomalla ilmaistu. Esim.

ÄITI JUOSTA EROTTAA LOPETTAA => Äiti juoksee (ja) repii (poikia) erilleen, "Lopettakaa!".

KATSOA ULKONA SATAA-LUNTA LUMI SATAA-LUNTA => (Hän) huomaa, (että) ulkona sataa lunta.

Adverbit ja adjektiivit: Erityisesti CA:ssa, mutta myös joidenkin verbaalien kohdalla tarkka käännösvastine edellyttää adverbien lisäämistä käännökseen. CA:ssa esimerkiksi ilmeellä ilmaistaan asennetta, olotilaa, tapaa jne. Joidenkin verbaalien kohdalla liikkeen modifioinnilla ilmaistaan tekemisen tapaa. Samoin on toisinaan adjektiivien kanssa. Suomenkieliset vastineet kirjoitetaan sulkuihin.

Esim: (ei-manuaalisesti "allapäin ilme") LÄHTEÄ(L) => (allapäin) (hän) lähtee-pois.

Yleisesti sulkujen käytöstä:

Sulkujen sisään kirjoitetaan ilmaukset, jotka

- 1 on ilmaistu pelkästään nonmanuaalisesti, (adverbimäiset ilmaukset, konjunktiot, tekijä, jne.) sekä
- 2) lisäykset, jotka on kirjoitettu, jotta käännösteksti olisi luettavaa (esim. kontekstista saatu informaatio, konjunktiot jne.).

Sulkeiden sisään merkitään pääsääntöisesti kokonaisia sanoja. Esimerkiksi suomen kielen edellyttämiä persoonapäätteitä yms. ei suluteta. Jos sulkujen käytön syy on useamman peräkkäisen sanan kohdalla sama, voi kaikki sanat laittaa samojen sulkujen sisään.

Lähteet

- ELAN. < http://www.lat-mpi.eu/tools/elan/elan-description_ > [viitattu 23.4.2018].
- FinSL Signbank 2018. Suomen viittomakielten korpusprojektiin tarkoitettu leksikaalinen tietokanta. < <http://signbank.csc.fi> > [viitattu 23.4.2018].
- Jantunen, Tommi 2003. Johdatus suomalaisen viittomakielen rakenteeseen. Helsinki: Finn Lectura.
- Jantunen, Tommi 2013. Signs and transitions: Do they differ phonetically and does it matter? *Sign Language Studies* 13:2, 211-237.
- Jantunen, Tommi 2014. Viittoman alfa ja omega annotaation näkökulmasta. Teknologisen kehityksen vaikutus suomalaisten viittomakielten tutkimukseen. Työpaja XLI kielitieteen päivillä Turussa 9.5.2014
< http://users.jyu.fi/~tojantun/ProGram/ProGram_files/41ktp_presentation.pdf > [viitattu 23.4.2018].
- Jantunen, Tommi 2015. How long is the sign? *Linguistics* 53:1, 93-124.
- Johnston, Trevor 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, Trevor 2016. Auslan Corpus Annotation Guidelines. Centre for Language Sciences, Department of Linguistics, Macquarie University, Sydney, Australia.
- Keränen, Jarkko, Syrjälä, Henna, Salonen, Juhana & Takkinen, Ritva (2016). The Usability of the Annotation. In E. Efthimiou, F. Stavroula-Evita, T. Hanke, J. Hochgesang, J. Kristoffersen & J. Mesch (Eds.), *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining* [organized as a part of LREC'16 at Portoroz, Slovenia, May 28, 2016]. Paris: European Language Resources Association (ELRA), pp. 111-116.
- Keränen, Jarkko, Nieminen, Henri, Salonen, Juhana, Green, Olga, Pippuri, Outi, Wainio, Tuija & Takkinen, Ritva 2018. C-FINSL:n ELAN-ohjeet. Tähän [www-linkki](http://www.cfinl.fi), josta ohjeet löytyvät
- Rissanen, Terhi 1985. Viittomakielen perusrakenne. Helsingin yliopiston yleisen kielitieteen laitoksen julkaisuja 12.
- Salonen, Juhana, Takkinen, Ritva, Puupponen, Anna, Nieminen, Henri & Pippuri, Outi (2016). Creating Corpora of Finland's Sign Languages. In E. Efthimiou, F. Stavroula-Evita, T. Hanke, J. Hochgesang, J. Kristoffersen & J. Mesch (Eds.), *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining* [organized as a part of LREC'16 at Portoroz, Slovenia, May 28, 2016]. Paris: European Language Resources Association (ELRA), pp. 179-184.
- Suomen viittomakielten korpusprojekti 2018. Jyväskylän yliopisto, Kieli- ja viestintätieteiden laitos, viittomakielen keskus.
<<https://www.jyu.fi/hytk/fi/laitokset/kivi/opiskelu/oppiaineet/viittomakieli/tutkimus/menoissa-olevat-projektit/suomen-viittomakielten-korpusprojekti>> [viitattu 23.4.2018].
- Suvi Suomalaisen viittomakielen verkkosanakirja. Kuurojen Liitto ry:n sanakirjatyo.
< <http://suvi.viittomat.net> > [viitattu 23.4.2018].
- Takkinen, R. 2008. Kuvailevat verbit suomalaisessa viittomakielessä. *Puhe ja kieli* 28, 17–40
- Wallin & Mesch 2018. Annoteringskonventioner för teckenspråkstexter. Version 7. Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet.