European Commission | Horizon 2020 European Union funding for Research & Innovation

Project Number: 693349

State of the art (Psycholinguistics)

Authors

Carlo Cecchetto, Jordina Sánchez Amat, Naama Friedmann, Josep Quer, Caterina Donati

Version 1.1 – 11/10/2017

Project title: The SIGN-HUB: preserving, researching and fostering the linguistic, historical and cultural heritage of European Deaf signing communities with an integral resource
Lead contractor: Universitat Pompeu Fabra
Contact person: Josep Quer Departament de Traducció i Ciències del Llenguatge Roc Boronat, 138 08018 Barcelona Spain Tel. +34-93-542-11-36 Fax. +34-93-542-16-17 E-mail: josep.quer@upf.edu
Work package: WP2
Affected tasks: Task 2.3

Nature of deliverable[1]	R	DEM	DEC	O
Dissemination level[2]	PU	PP	RE	CO

[1] *R: Report, DEM: Demonstrator, pilot, prototype, DEC: Websites, patent fillings, videos, etc., O: Other*

[2] ***PU**: public, **PP:** Restricted to other programme participants (including the commission services), **RE** Restricted to a group specified by the consortium (including the Commission services), **CO** Confidential, only for members of the consortium (Including the Commission services)*

COPYRIGHT

European Commission | Horizon 2020 European Union funding for Research & Innovation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 693349.

History of changes

VERSION	DATE	CHANGE	REVIEWER(S)
1.0	31/03/2017	Initial version.	Carlo Cecchetto, Jordina Sánchez Amat, Naama Friedmann, Josep Quer, Caterina Donati
1.1	29/09/2017	History of changes added. Section 7 updated.	Caterina Donati, Carlo Cecchetto, Jordina Sánchez Amat

CONTENTS

1. Scope of the document 5

2. Introduction 6

3. Procedures in sign language assessment tools 7

4. Features analyzed in sign language assessment tools 10

5. Implications of lexical processing in the development of assessment tools 14

6. Considerations for the development of syntactic assessment tools 19

6.1. Tests to assess poor- and high-performance 19

6.2. Research on language disorders in deaf signers 20

6.3. Research on first language late exposure 22

7. Final remarks 24

8. References 26

1. Scope of the document

In this document, we aim to review existing sign language assessment tools, their strengths and weaknesses, and some results obtained from them. This discussion is crucial for the development of the assessment tools planned as Task 2.3 of SIGN-HUB.

2. Introduction

Haug (n.d., 2008, 2011a, 2011b) provided the most extensive contribution to the knowledge and discussion about the development of sign language assessment tools, although other reviews have already been published that describe these tools and critically discuss procedures, normative samples and test formats (Enns et al., 2016; Lieberman & Mayberry, 2015; Mann & Haug, 2014; C. R. Marshall & Morgan, 2015; Singleton & Supalla, 2003). In this document, we aim to contribute to this debate by analyzing specific tasks to assess linguistic features, their strengths and weaknesses. We consider the different aims the assessment tools are developed for: studying sign language acquisition, analyzing the effects of late first language acquisition, investigating sign language disorders (aphasia, specific language impairments, etc.), checking relationships between sign language and literacy and/or spoken languages, etc. Although we will focus on sign language assessment tools created to measure sign language proficiency, we will also take into account some specific instruments used in experiments, since we consider them useful for the discussion on the development of SIGN-HUB assessment tools. Having said that, our review cannot be fully comprehensive but mostly focus on discussing tools (or part of them) that can give us more direct indications on tests that will be designed as part of Task 2.3.

3. Procedures in sign language assessment tools

Procedures to administer tests vary from naturalistic observation to very specific experimental tasks, designed to test the mastery of specific language features. The *American Sign Language-Proficiency Assessment* (ASL-PA) (Maller, Singleton, Supalla, & Wix, 1999) uses interviews and peer interaction observations to check the presence of the eight linguistic structures identified for the instrument. In the *MacArthur Communicative Developmental Inventories* adapted into sign languages (Anderson & Reilly, 2002; Woolfe, Herman, Roy, & Woll, 2010; cf. https://volis.it/ for an adaption to LIS) and other assessment tools based on parental and/or teacher reporting (Lichtig et al., 2011; Simms, Baker, & Clark, 2013)[3] naturalistic observations are used to analyze the production or comprehension of specific vocabulary and/or the child's communicative skills.

Some tools to assess communicative and linguistic behavior are not exclusively devised to assess sign language, but any communicative skills in both signed and spoken languages, as the *Profile of Multiple Language Proficiencies* (Goldstein & Bebko, 2003) and the protocol evaluating communicative abilities developed by Lichtig et al. (2011).

Play-based assessment is used in some tests, as part of the communicative skills assessment tools by Lichtig et al. (2011). Another similar approach, although less naturalistic, is dramatization with dolls of signed texts, as in a task of the *Aachen Test for Basic DGS Competence*[4]: after viewing an explanation in sign language, participants are asked to reproduce it by means of dolls in order to check comprehension.

Description of images, real objects or filmed events, is used in a number of tests. Several tasks of the *Test Batteries for ASL and Auslan Morphology and Syntax* (Schembri et al., 2002)[5] show filmed events and ask the participant to describe them. Responses are videotaped and analyzed afterwards. The *American Sign Language Assessment Instrument* (ASLAI) (Hoffmeister et al., 2015) aims also to assess the mastery to describe images but instead of recording the participant's answer, 4 possible descriptions in ASL are presented. The participant is asked to select the best option. A simpler task involving images is the picture-naming task, which is the case of part of the *British Sign Language Receptive Skills Test* (Herman, 2002), and of the assessment instrument for Sign Language of the Netherlands –Nederlandse Gebarentaal, NGT– (Hermans, Knoors, & Verhoeven, 2010).

Pictures are also widely used in comprehension tasks, as in sign-picture matching tasks of the assessment instrument for NGT (Hermans et al., 2010), the *American Sign Language Comprehension Test* (ASL-CT) (Hauser et al., 2016), and the *Test Batteries for ASL and Auslan Morphology and Syntax* (Schembri et al., 2002). This procedure is also used at the sentence level: sentence-picture matching is present in the *Test Batteries for ASL and Auslan Morphology and Syntax*(Schembri et al., 2002), the receptive morpho-syntactic tasks of the assessment instrument for NGT (Hermans et al., 2010), the Computer-based Test for German Sign Language (CTDGS) (Mann, 2006, 2007) and in the BSL-RST (Herman, 2002) and its adaptations to German Sign Language –DGS– (Haug, 2011a), ASL (Enns & Herman, 2011), Spanish Sign Language –LSE–

[3] The Signed Language Development Checklist by Judith Mounty is also based on live observation and recorded language samples (Haug, n.d.). However, materials on this instrument are no longer available (J. Monty, personal communication, March 21, 2017).

[4] Test developed by the Rheinisch-Westfälische Technische Hochschule, http://www.signges.rwth-aachen.de/1717.html?&L=xwumorzwlhogb (Haug, n.d.).

[5] The Test Battery for ASL Morphology and Syntax description was not available, and we followed Schembri et al (2002) description of both tests for ASL and Auslan (the article explains the adaptation of the battery from ASL into Auslan).

(Valmaseda, Pérez, Herman, & Ramírez, 2013), Auslan (Johnston, 2004) and Polish Sign Language (Kotowicz et al., 2016).[6]

The inverse process (picture to sentence matching) is used in several tasks of the ASLAI, the *Real Objects and Plurals* task and the *Syntax-simple task* (Hoffmeister et al., 2015): a picture or a real world event is displayed and the participant must select, among 4 sentences in ASL, the one that describes the image or video.

Repetition tasks are used to test sublexical features as phonological processing. That is the case of the *Nonsense Sign Repetition Task for British Sign Language* (Mann, Marshall, Mason, & Morgan, 2010) and the assessment instrument for NGT (Hermans et al., 2010). As for sentence-level tasks, repetition is also present in the *American Sign Language-Sentence Reproduction Test* –ASL-SRT– (Hauser, Paludneviciene, Supalla, & Bavelier, 2008; Supalla, Hauser, & Bavelier, 2014),[7] in the SignMET assessment tools (SignMET, 2013), [8] and for BSL (C. R. Marshall, Mason, et al., 2015).[9] Mayberry & Eichen (1991) and Mayberry and Fischer (1989) also used this method on their research on effects of first language late exposure. This kind of task measures both receptive and expressive skills (Quinto-Pozos, Singleton, & Hauser, 2016).

Narratives are used to assess content comprehension as well as to elicit data to analyze grammatical and narrative aspects. As for narrative comprehension, participants view a story and several questions are raised, either literal or inferential, which is the case of ASLAI (Hoffmeister et al., 2015), the CTDGS (Mann, 2006, 2007) and the assessment instrument for (Hermans et al., 2010). In this line, the *VOLIS test –Valutazione online della LIS–* provides three answers to each question, and the participant must select the right one (Istituto di Scienze e Tecnologie della Cognizione - Consiglio Nazionale delle Ricerche, n.d.).

Narrative production is elicited through story retelling, after viewing a cartoon movie or a wordless picture book –both stimuli are used in the *Test of American Sign Language* (Strong & Prinz, 1997)–. The (silent) cartoon movie stimuli is used in the BSL Narrative Production Test (BSL-PT),[10] the SignMET tools (SignMET, 2013) and the *Test Batteries for ASL and Auslan Morphology and Syntax* (Schembri et al., 2002), whereas the wordless picture book is used in the assessment instrument for NGT (Hermans et al., 2010).

Finally, we identified the use of grammatical judgements in three different tests: the *Grammatical Judgment Test for ASL* (Boudreault & Mayberry, 2006), as well as its adaptation into BSL (Cormier, Schembri, Vinson, & Orfanidou, 2012) and the *VOLIS test* (Istituto di Scienze e Tecnologie della Cognizione - Consiglio Nazionale delle Ricerche, n.d.). In these tests, a sentence is shown that may contain grammatical violations or not. Participants are asked to press

[6] No sources about the adaptations cited by Haug (n.d.) into Japanese Sign Language and Danish Sign Language could be found. Adaptation into Italian Sign Language could never be completed due to technical problems (L. Surian, personal communication, March 23, 2017). It has also been adapted into NGT (Klatter-Folmer, van Hout, Kolen, & Verhoeven, 2006), but the article does not go into detail about the adaptation process.

[7] The ASL-SRT was being adapted into DGS in 2015 (Hauser, Quinto-Pozos, & Singleton, 2015), and has been adapted into BSL (Denmark & Atkinson, 2015).

[8] The SignMET developed assessment tools for Catalan Sign Language (LSC), Italian Sign Language (LIS), Swiss German Sign Language (DSGS) and French Sign Language (LSF).

[9] The BSL Sentence Reproduction Test used by Marshall, Masion, et al. (2015) was designed before the adaptation of the ASL-SRT into BSL took place (K. Rowley, personal communication, March 27, 2017).

[10] The BSL Narrative Production Test is described by Marshall, Jones, et al. (2015). Originally it was called BSL Production Test (Jones et al., 2016). The reference of the tool is the following:

Herman, R., Grove, N., Holmes, S., Morgan, G., Sutherland, H., and Woll, B. (2004). Assessing BSL Development: Production Test (Narrative Skills). London: City University.

This test has been adapted into Auslan, and an adaption into ASL and LSE is planned (Enns et al., 2016).

one of two buttons if the sentence is identified as 'correct' or 'incorrect'. On the other hand, the ASLAI Syntax-Difficult Task displays four versions of the same sentence, and participants are asked to select the 'correct' one.

4. Features analyzed in sign language assessment tools

A wide range of features are studied in the various sign language assessment tools listed in the preceding session: communicative skills, phonological processing, lexical knowledge, linguistic structures and narrative proficiency. Comprehension and/or production skills may be assessed in each of these domains. Depending on the aims of the instrument, it focuses on one or several areas and the way to administer it differs.

Phonology receptive tasks test whether the signer is able to distinguish between two signs that form a minimal pair, as does the assessment instrument for NGT (Hermans et al., 2010). This test also measures expressive phonological skills through an imitation task: participants are asked to repeat a sign viewed previously on a screen, and the test administrator evaluates the accuracy for a specific parameter. In the Lexical Phonology Production test of the *Test Battery for Australian Sign Language Morphology and Syntax* participants "are asked to produce the citation form of particular signs. The signs elicited by the test include handshapes and movements that resemble components of the morphologically complex signs elicited by latter tests (such as the Verbs of Motion Production test)" (Schembri et al., 2002, p. 22). Phonological judgment task is used in the *ASL phonological awareness* (ASL-PA). Four images are shown to the participant: a cue, the target response which shares one or more phonological parameters with the cue, and two distractors (McQuarrie & Abbott, 2013).

Lexical knowledge and/or processing is mainly assessed through picture-naming tasks (expressive vocabulary) and sign-picture matching tasks (receptive vocabulary). However, some tests include a sign meaning task, which consists in asking participants to define a sign, as in the assessment instrument for NGT (Hermans et al., 2010)(Hermans et al., 2010), and also to rate the correspondence between a sign and a picture, as in the Aachen Test for Basic DGS Competence (Haug, n.d.). In the ASLAI, several tasks are designed to assess lexical knowledge. The Vocabulary Difficult task (VOCD) presents a definition in ASL and the participant must choose, among four signs, the one that corresponds to the definition (Hoffmeister et al., 2015). To assess the antonyms and synonyms knowledge, ASLAI shows a sign and the participant is asked to select, among four choices, the ASL sign that has the same or the opposite meaning. Another task related to lexical knowledge is the ASLAI Vocabulary in Sentences: a sign is showed and the sentence where it fits the best among four choices must be selected.

Linguistic structures that are most commonly assessed in the tests that we were able to analyze are verb agreement, semantic classifiers, size/shape specifiers (or descriptive classifiers), number and role shift.

As for *agreement verbs*, the ASL-PA considers two features of this linguistic structure: real-world location(s) marking and abstract location(s) marking. If present in the participant's elicited data, each feature scores 1 out of 23.

Agreement verbs are also assessed in the BSL Narrative Production Test. The BSL grammar scores the following structures (total score = 30), "reflecting the number of different verb forms targeted" (Herman, Rowley, Mason, & Morgan, 2014, p. 348):

(i) spatial verbs including classifiers (12 points), e.g. PERSON-GO-TABLE
(ii) agreement verbs (4 points), e.g. SHE-GIVES- HIM
(iii) manner inflections (5 points), e.g. CHEW- HUNGRILY
(iv) aspectual inflections (5 points), e.g. TAKE- REPEATEDLY
(v) role shift (4 points), e.g. use of body orientation to identify each character.

However, there is a disadvantage when scoring the number of structures present in narratives: a signer summarizing the story may score low, although the few structures produced are grammatical. This led Sánchez Amat (2015) to suggest analyzing relative values (number of linguistic structures/number of clauses produced) instead of absolute values.[11]

In the assessment instrument for NGT, verb agreement was assessed in both the receptive and expressive morphosyntactic skills tasks, based on sentence-picture matching and picture description. The expressive morphosyntax task consists of 24 items. We were not able to find any information about the number of items in the receptive task. In both tasks, several morphosyntactic features are tested: verb agreement, aspect, classifier verbs of motion and location (Hermans et al., 2010, p. 110). Information about the number of items that belong to each feature was not found.

In the *Test Batteries for ASL and Auslan Morphology and Syntax* both comprehension and production of agreement verbs are assessed. Comprehension of agreement verbs is assessed by asking the participant to retell what a signer has said previously as if he or she is addressing to the signer. This was done in two different tasks: the first including items with only one sentence (e.g. GIVE 3sg to 1sg) and the second including two sentences (e.g. GIVE 3sg to 3sg, then GIVE 3sg to 1sg). Therefore, the participant needs to modify the verbs in order to maintain agreement. This procedure involves also production. Another task in the ASL battery assessing production is the Verb Agreement Production task, in which participants watch a series of film events performed by two actors, and are then required to sign the events they watched (Schembri et al., 2002).

Regarding agreement verbs production, in the assessment instrument for NGT they are analyzed by asking the participant to retell an event that has been shown to him/her on a video.

As a part of a project of building assessment tools for NGT, Jansma, Knoors & Baker (1997) developed a test including two tasks relevant for agreement verbs (Haug, 2008):

1) A receptive localization task- as localization is a pre-requisite for comprehension and production of verb agreement. The participant is asked where a specific referent is localized or which point in space is associated with a particular referent.
2) A receptive verb agreement task- participants watch a signed sentence and have to choose a matching picture out of three pictures.

Verb agreement was also assessed in grammaticality judgement tasks. The ungrammatical sentences were created by moving the verb (phrase) to a position where it cannot appear.[12] The word order alteration results in structures as the following (Cormier et al., 2012, p. 55):

(1) *Agreement verb sentence (grammatical)*
_____br
SIGN CLASS POSS-1 STUDENT++ REJECT-3
'My students couldn't be bothered with the sign class.'

(2) *Agreement verb sentence (ungrammatical)*
_____br
*SIGN REJECT-3 CLASS POSS-1 STUDENT++
'Sign reject class my students.'

[11] This work was a first attempt to develop tools to conduct cross-linguistic researches to compare narratives in Catalan Sign Language (LSC) and written Catalan or Spanish.

[12] In ERP studies violation with agreement verbs was produced by making the verb end "in an unspecified location opposite the location of the associated referent" (Hosemann, 2015, p. 85).

Semantic classifiers are analyzed in the ASL-PA under the Verbs of Motion feature. This category groups three structures (each one scoring 1 point if present in the child linguistic production): simple path movements (in absence of the handshape classifier), central object handshape classifiers, and secondary object handshape classifiers (the *ground*). These structures reflect the acquisition pattern according to Maller et al (1999). The *Test Batteries for ASL and Auslan Morphology and Syntax* (Schembri et al., 2002) also distinguishes between central and secondary object. In an image description task, 40 items involve only central objects, and 40 items involve both primary and secondary objects. In the assessment instrument of NGT, expression and comprehension of classifier verbs of motion and location were assessed by means of picture description and sentence-picture matching, respectively. The ASLAI Real Objects and Plurals task checks the skills on these constructions through real object/event-sentence matching. The Grammatical Judgment Test for ASL contained different types of classifiers. The authors state that it was the most difficult structure to make ungrammatical. Finally, they found only one way to make sentences ungrammatical: "switching the temporal order of the classifiers was the only violation that all judges agreed rendered the sentences unacceptable" (Boudreault & Mayberry, 2006, p. 619). Finally, the *BSL Receptive Sills Test* also assesses this structure. According to the list of items for DGS-RST by Haug (2011a, pp. 293–299), the linguistic feature spatial verb morphology groups classifier constructions and lexical verbs (with location agreement). This category is quite unspecific: it does not distinguish between classifier constructions and lexical verbs, and almost half of the items belong to it in the DGS version (21 items out of 50). This unspecificity may hamper interpretation of results.

Size/shape specifiers, or *descriptive classifiers*,[13] is a linguistic feature assessed by the BSL-RST and its adaptations. However, only 3 out of 50 items belong to this feature (at least in the DGS version). The target and distractors involved are as follow (Haug, 2011a, p. 295):

(3) *Target*: CURLY-HAIR (image: person with long curly ringlets)
Distractors: long wavy hair; short straight hair; long frizzy hair

The *Test of ASL* has a specific task to assess classifier comprehension. According to the description provided by Niederberger (2004) in her adaptation of the TASL into LSF, 8 out of 10 items involved descriptive classifiers, but there are also semantic classifiers. A picture-sentence matching task was used. Therefore, we consider that it is closer to a production task: the participant is asked to select the best way to describe the image previously displayed. Therefore, comprehension is taken for granted, since the picture offers the concept which is being described. When asking to select the ASL sentence that matches the picture, one is checking if the participant knows what the best way to describe the image is. Descriptive classifiers are assessed in the classifiers task of the *Grammatical Judgment Test for ASL*, in the same way described above for the semantic classifiers task.

Number is usually assessed together with two other features: either aspect or spatial arrangement/distribution. The ASL-PA, the assessment instrument for NGT and the *Test Batteries for ASL and Auslan Morphology and Syntax* consider number related to aspect. Expression of distribution or the number of times an action happens is assessed. On the other hand, ASLAI, TASL and BSL-RST focus on spatial arrangement. As expression of number often involves semantic classifiers, information about spatial arrangement is provided by means of classifier constructions as well.

Regarding *role shift*, in the *ASL Receptive Skills Test* it was one of the two new linguistic features –along with conditionals– added with respect to the original *BSL-RST* (Enns & Herman, 2011). This decision was made with only the six original categories in which 10-year-old participants reached ceiling performance. Although some targets of the *BSL-RST* with role shift were already present, they were not specifically assessed. Role shift is also analysed in elicited linguistic data with the *ASL-PA* (Maller et al., 1999) and is a score sheet item of the *Test of ASL* (Strong

[13] This categorization is still under discussion (Zwitserlood, 2012).

& Prinz, 1997) and its adaptations. Niederberger's adaptation considers scoring one to three occurrences of role shift with score 1 (in a 0-2 scale) and score 2 for three occurrences very well developed or more than three (Niederberger, 2004). In a similar way, BSL-PT rate from 0 to 4 mastery of role shift (Enns et al., 2016). In addition, role shift is considered in the SignMET comprehension task as a feature needed to be understood by participants in order to be able to answer the content questions (SignMET, 2013).

Other linguistic features assessed are: pointing, negation, syntax (word order and relative clause sentences), noun-verb pairs, pronouns, conditionals and wh-sentences.

At the **discourse level**, the domains that have received most attention are discourse cohesion and narrative coherence.

As for *discourse cohesion*, the *Test of LSF* assesses discourse cohesion markers (Niederberger, 2004, p. 296). Animate referents in discourse are categorized depending on their referential function: introduction, maintenance and reintroduction. This approach, which follows in some way Givón's framework (Givón, 1984), has been adopted in several researches on Deaf signers' narratives (Cormier, Smith, & Zwets, 2013; Frederiksen & Mayberry, 2014; Morgan, 1998; Reynolds, 2016; Sánchez Amat, 2015) as well as in sign language L2 learners (Bel, Ortells, & Morgan, 2014; Frederiksen & Mayberry, 2015).

Still in the field of referential cohesion, the *Computer-based Test of DGS* (Mann, 2006, 2007) aims at assessing *referential distinction*, which "provides the language learner with the information needed in order to distinguish among the thematic roles and grammatical functions of different elements (i.e. agent, patient) within a sentence" (Mann, 2007, p. 62). It is a construct which serves the specific purpose of comparing mastery in this domain both in DGS and in German in different conditions (picture-aided, context-based and translation from one language into the other). To the best of our knowledge, this tool does not distinguish between different structures (the various referential forms) intervening in referential distinction. This test assesses referential distinction in the sentence and the discourse levels (in two different tasks).

Narrative production skills in sign language have been assessed from two different perspectives: the story grammar (Stein & Glenn, 1975) and the causal network (Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985) approaches. The former has been adopted (and simplified) in the TASL and the TLSF, which analyze the presence of the setting (presentation of characters, place, time and initiating event), episodes and episodic relations, and resolution (Niederberger, 2004, p. 295). A similar approach is used for the BSL Narrative Production Test, which adopt the high point analysis by Labov. From this perspective, "narratives are coded for orientation; complicating actions; climax; resolution; evaluation and sequence" (Enns et al., 2016). Vercaingne-Menard, Godard, & Labelle (2001) also used this approach in their longitudinal study of narrative skills of two Deaf signer children. The latter perspective, the causal network model, has been used by Marschark, Mouradian, & Halas (1994) and Sánchez Amat (2015). Episodes are conceived as goal-attempt-outcome (GOA) structures. Setting and psychological relations (cognitive or emotional responses/motivations of the characters) are also analyzed.

The assessment instrument for NGT also assesses narrative production skills, but no information has been found about the scoring methods using in this part.

Narrative comprehension skills are assessed in several tools by means of content questions after viewing a story in sign language (see the previous section for a list of tests that include this kind of task).

5. Implications of lexical processing in the development of assessment tools

The preparatory work for the creation of Lexical tests for SIGN-HUB started from a general reflection about the nature of the sign language lexicon.

In sign languages, a distinction may be drawn between the core (or frozen) lexicon and the non-core (or productive) lexicon. The core lexicon refers to the lexicon proper: those word/sign forms which are a manifestation of a given lexeme. However, in addition to the core lexicon, nearly all sign languages make use of mechanisms which involve the combination of phoneme-like units (like handshapes, locations, movements, etc.) with a more gestural component. This is the case of classifiers. As they often identify a class of objects by visually representing some properties that these objects share, i.e. their size, shape, or the way they are handled, classifiers tend to be iconic to some extent.

A series of properties characterize and differentiate the core and non-core lexicon (we base this short summary on Göksel & Pfau, in press): the core lexicon is subject to phonological constraints while the non-core lexicon frequently violates them. In the non-core lexicon, space and movement are often used topographically/isomorphically. Finally, core lexicon items can belong to any of the traditional parts of speech while classifiers are mostly used in predicative constructions.

Importantly, in some cases the distinction between core and non-core lexicon is blurred, either because a classifier is on its way to becoming a fully conventional sign or because some signs belonging to the core lexicon can delexicalize, when the location/orientation of the sign becomes isomorphic with the location/orientation of the referent (this can happen only if the referent is concrete and the sign is not articulated on the body).

Faced with the distinction between core and non-core lexicon, researchers in Task 2.3 need to decide which items to use in lexical tests. Also, based on the study of the previous literature, which we are about to summarize, the decision is favoring items that belong to the core lexicon. We opt for this decision, not because we underestimate the importance of non-core lexicon in sign languages, but mainly for two reasons.

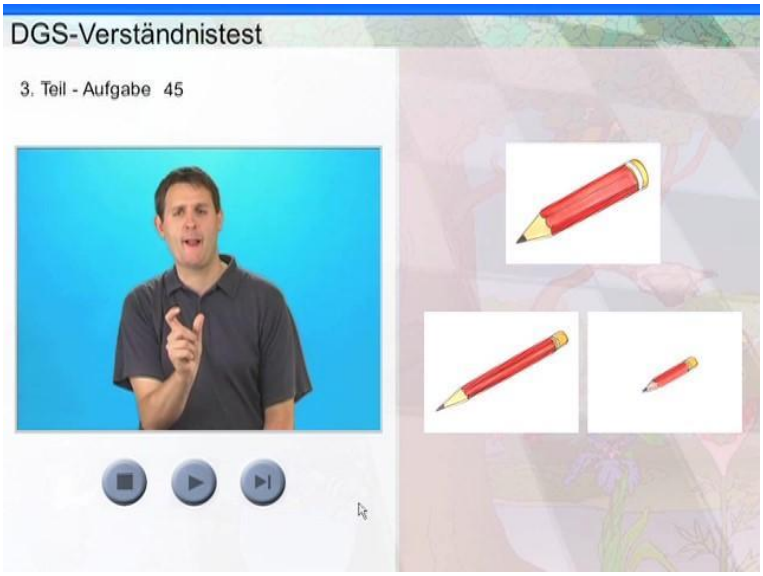
The first one is that, given the fact that the core lexicon is used in predicative constructions (classifier constructions), it is very difficult to test the non-core lexicon competence of a signer with a sign in isolation. The non-core lexicon is inherently clausal, so an assessment of its correct mastery is bound to be a syntactic (rather than) a lexical test.

The second reason is more practical and has been suggested by a review of the existing literature on lexical assessment in sign languages: as classifiers tend to be iconic, they are unsuited for picture-matching tasks, which are the most common procedures used in assessment tests.

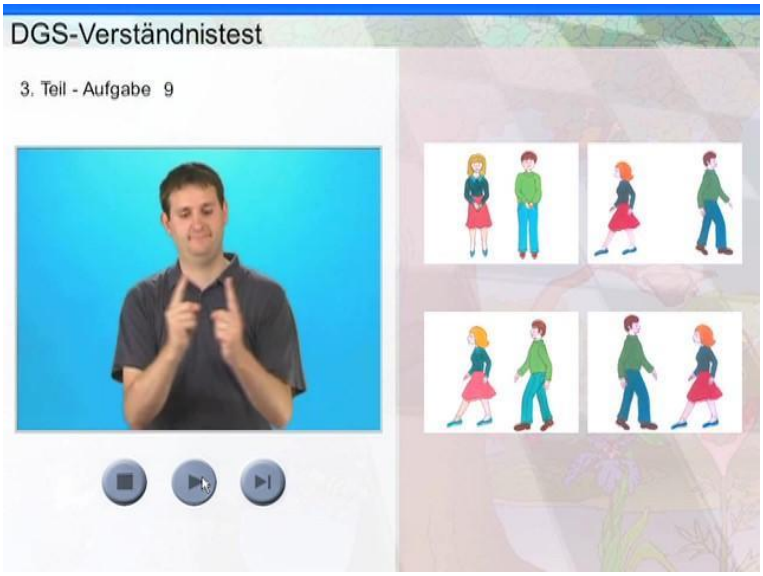
For these reasons, the tests we are designing for SIGN-HUB will be different from the most established test of sign language assessment, namely the BSL Receptive Skills Test (Herman, Holmes, & Woll, 1999). This test may be taken as a reference point in the sign language assessment test literature, since it has been adapted (or partially adapted) to many sign languages including Danish Sign Language, Australian Sign Language, German Sign Language, Japanese Sign Language, Spanish Sign Language, Polish Sign Language, American Sign Language (source: http://www.signlang-assessment.info/index.php/british-sign-language-receptive-skills-test.html) and Swiss German Sign language (Enns et al., 2016).

The Receptive Skills Test has been initially conceived for children aged 3-13. They are shown a video and respond by selecting the most appropriate picture from a choice of three or four. The lexical items that are used in the Receptive Skills Test do not need to be part of the core lexicon and often do not belong to it, since an explicit goal of this test is assessing the correct

use of size and shape and handling classifiers and more generally the ability to recognize the topographical use of space. The result is that there is a potential risk that the participant can make a correct guess based on the iconic correspondence between the classifier and the depicted object. As an example, we provide two items from the adaption to DGS of the Receptive Skills Test (Haug, 2011, full videos available at http://www.signlang-assessment.info/index.php/german-sign-language-receptive-skills-test.html). We suspect that in the item below an adult naïve to sign languages might give the correct answer ("the short thin pencil") based on the limited extension of the handshape and on the presence of sucked-in cheeks.



For the following item, we suspect that a person naïve to sign languages might give the correct answer (the picture in the lower left box) if the participant notes that the collocation of the classifiers for person in the signing space mimics the collocation of the two people facing each other in the drawing.



We acknowledge that the risks that participants respond correctly by using the heuristics that are in principle available to non-signers is minimized with small children. We also recognize that measuring these heuristics in children might be an indirect probe for their linguistic capacities, since these heuristics are likely to be boosted by sign language use. However, as the tar-

get population of SIGN-HUB includes adults, we think that a using non-core lexicon in lexical assessment tests should be avoided as much as possible.

Having said that, iconicity is a component of sign languages that cannot be artificially eliminated and many signs that belong to the core lexicon class have an iconic origin that they may retain to various degrees. More generally, a component of iconicity is not incompatible with Saussurian arbitrariness.

Two lexical tests that are more based on core lexicon and are confronted with this problem, are the British Sign Language Vocabulary Test (Mann & Marshall, 2012) and Perlesko (*Prüfverfahren zur Erfassung lexikalisch-semantischer Kompetenz*), a receptive vocabulary test for DGS (cf. Bizer & Karl, 2002). Information on Perlesko is available in English at http://www.signlang-assessment.info/index.php/perlesko-vocabulary-test-for-german-sign-language.html.

We focus here on the British Sign Language Vocabulary Test because Mann & Marshall (2012) discuss methodological choices and problems that had to be dealt with during test construction, so it is particularly useful for our work in SIGN-HUB.

The British Sign Language Vocabulary Test was developed for children between the ages of 5 and 15 years. The same 120 items belonging to the core BSL lexicon (nouns, verbs, and adjectives) were used in the four different tasks that compose this test.

- TASK 1: this is a receptive task (sign-to-picture) in which participants are shown the target BSL sign, followed by four pictures, and have to select the picture that corresponds to the target sign.
- TASK 2: in this receptive task (picture-to-sign), participants are shown a picture, followed by four BSL signs, and have to select the sign that matches the picture. For both receptive tasks, one of possible responses is of course the correct one, the second one is a phonological distractor, the third one is a semantic distractor, and the fourth one is distractor with no phonological or semantic relation with the target sign.

The choice of core lexicon items minimizes, but does not eliminate, the possibility of guessing based on pure iconicity.[14] Mann & Marshall (2012) handle this issue (i) by excluding items that have a high level of iconicity (ii) by choosing *distractors* that to a non-signer eye might be in an iconic relation with the target sign. This point can be better explained by making a concrete example. A non-signer in principle might associate the sign for DOG in BSL to the picture of a dog because the sign refers to the dog's ears.

[14] Iconicity did not affect results in a naming task conducted with a deaf user of BSL with anomia, 'Charles' (J. Marshall, Atkinson, Smulovitch, Thacker, & Woll, 2004), whereas his gesture system was intact after the CVA. Atkinson, Marshall, Woll, & Thacker (2005) found similar results in their research with 15 participants with brain damage. These findings would suggest that the role of iconicity in naming tasks could be less crucial as hypothesized until now in people with aphasia, and more in general, "this suggests that the neural organisation of the BSL lexicon does not distinguish iconic from non- iconic signs." (Atkinson et al., 2005, p. 246) However, more research is needed on this topic.

The sign for DOG in BSL (souce: https://www.spreadthesign.com/)

However, the sign DOG in BSL might also refer to a knife and fork being held for eating. Therefore, a distractor for the sign DOG in TASK 1 was the image of cutlery, the rationale being that a guess based on pure iconicity could lead a non-signer to choose this picture. Only knowledge of the BSL lexicon guarantees that the picture of the dog is preferred to the picture of cutlery.

The British Sign Language Vocabulary Test includes also two production tasks:

- TASK 3: children see a picture and are asked to produce the corresponding sign.
- TASK 4: children see a sign and have to generate another sign with an associated meaning (say, they see the sign DOG and are expected to produce the sign BONE). In later version of TASK 4, the participant needs to generate three associated signs.

The results reported by Mann & Marshall (2012) indicate that TASK 1 is the easiest task, followed by TASK 2, with TASK 3 and TASK 4 being the hardest.

The psycholinguistic literature on sign languages discusses other tests that presuppose lexical knowledge but are complex because they involve the manipulation and/or retention of lexical items. These tests are not the focus of SIGN-HUB task 2.3. but are mentioned here and their possible use with special population is briefly discussed.

Several papers focus on short-term memory for list of signs. In particular, a number of studies on ASL, including Wilson & Emmorey (1997), Wilson & Emmorey (1998) and Bellugi, Klima & Siple (1975) show that that the architecture of short-term memory for signs closely resembles the architecture proposed for spoken languages. This suggests that phonological short-term memory is not an exclusive domain of the auditory modality but emerges as an a-modal mechanism, triggered by language in any modality. However, an important difference emerges, namely the sign span is smaller than the word span (span is defined as the longest sequence in which serial recall is correct). The difference has been replicated with different populations (deaf signers compared to hearing speakers and hearing bilinguals tested with both sign and spoken language material) and with several sign languages. As an example of the span difference, Geraci, Gozzi, Papagno & Cecchetto (2008) report that the mean span for auditorily presented words was 4.94 while the mean sign span was 3.31 (the sequences of words were the translation of the sequences of LIS signs and words and signs were matched for time of articulation). Given that the sign span may be as low as 3 in the general population of signers, it may be difficult to identify short-term memory deficit for signs by using a simple span task. Free recall is less studied in sign languages, but recent work by Alba (2016) replicates the word advantage. All in all, the lesson that emerges is that a low performance in recall tasks involving signs should not be mistaken for pathological, unless more specific tests are built.

A way to probe the lexical organization, which is commonly used for spoken languages, is semantic and phonological fluency tasks. The tasks require participants to produce as many words as they can in a given time (usually a minute) that fall into a certain semantic category (usually "animals" or "food") or that begin with a certain letter. Marshall, Rowley & Atkinson (2014) adapted these tasks to BSL. While the semantic fluency task can be administered in the same way as in spoken languages, the phonological fluency task requires some adaptation for obvious reasons. Marshall and colleagues involved a deaf signer to demonstrate a particular handshape, location and movement and asked participants to produce as many signs as they could with that parameter value. The results differentiate the two fluency tasks. Comparing the number of correct responses made by signers to "animals" with results from adults of a similar age range in four spoken languages reveals that users of BSL produce a comparable number of responses. Comparing the phonological fluency results to those of spoken languages is more complex, because of differences in how the task is administered. However, a comparison to spoken languages reveals the relative difficulty of phonological categories for signers. Marshall and colleagues report that while in spoken language phonological fluency falls in the region of 10–15 words, in BSL the "above the shoulders" location elicits 19.53 signs whereas other categories, e.g. the "I" handshape and "palm of the non-dominant hand" (M=7.40 and 8.13, respectively), elicit a small number of signs.

Based on this research, one might conclude that the fluency tasks might be used with a pathological population, with the proviso that the phonological fluency task requires prior identification of a specific parameter value which is shared by many signs in the given language (as the "above the shoulders" location plausibly is).

Finally, Mayberry & Witcher (2005) used a priming task to probe lexical organization in ASL. The stimuli were prime-target sign pairs. Half of the targets were existing signs, which stood in one of three linguistic relationships to the primes: some of the targets formed a minimal phonological pair with the prime, others were semantically related to the prime, and the remaining signs shared no linguistic relation to the target. The other half of the targets were non-signs, created by altering one parameter of an ASL sign. The participant decided with a button press whether the target is a real sign or not. Results indicate that age of first exposure affected response time to recognize signs. Specifically, as age of first exposure increased, reaction times also increased. Furthermore, substantial phonological overlap between the prime and target facilitated sign recognition for the early learners, but inhibited it for the late learners (there was no facilitation with native learners, possibly due to a ceiling effect). By contrast, semantic overlap between prime and target facilitated sign recognition for all the groups independently of age of first exposure.

These results (especially the inhibitory effect of a phonologically-similar prime for late learners) suggest that tasks involving phonological priming might be sensitive to lexical impairment.

6. Considerations for the development of syntactic assessment tools

In the field of syntactic skills, sign language assessment tools vary on whether they are designed to measure mastery of specific features or their aim is a general assessment of sign language proficiency.

In this section, we will discuss advantages and disadvantages of assessment tools depending on whether they assess specific linguistic features.

6.1. Tests to assess poor- and high-performance

The DGS-RST and its multiple adaptations and the ASL Sentence Reproduction Test are tools designed to get a general measure of proficiency in sign language, to the best of our knowledge. As discussed above (see section §4, Features analyzed in sign language assessment tools) for some features there are few items –for instance, there are only three items for size/shape specifiers in the DGS-RST– whereas other features are much more represented –this is the case of the 'spatial verb morphology' feature, with 21 items in the DGS-RST. This seems to be also the case for the assessment instrument for NGT, but as said above (see footnote 6) the information available is not enough to confirm it.

The ASL Sentence Reproduction Test also identifies the syntactic complexity of the sentences used in the test, but the categorization is quite generic: the simpler sentences are transitive and adjectival predications, whereas the more complex sentences are combinations of locative predications, transitive predications, adjectival predications, etc., and are morphologically more complex (Hauser et al., 2008). However, the specific grammatical features that make sentences more complex are not taken into consideration. This is why researchers need to complement the test with an analysis of errors: "while the test can function as a screening instrument for overall proficiency levels, its full value as a diagnostic instrument is realized with additional error analysis of responses" (Supalla et al., 2014, p. 14). The error type analysis lead them to hypothesize that highly fluent signers can "retain the deep structure of target sentences, whereas less fluent signers made a variety of more superficial errors and changes". The fact that the highly fluent participants process the semantic content fast leads them to re-construct the sentence less literally. The researchers hypothesize that their results may be in line with those obtained with late learners by Mayberry and Fischer (1989), but they need to deepen on the analysis of the data to obtain more support. Mayberry and Fischer (1989) asked participants (deaf native and non-native signers) to watch and reproduce simultaneously ASL narratives in good and poor viewing conditions. They focused on the analysis of lexical changes of two kinds: semantic and phonological changes.

On the other hand, The Test Batteries for ASL and Auslan Morphology and Syntax do investigate specific sign language features. Unfortunately, few studies have been conducted using these tools,[15] and no investigation has been conducted with population with language impairments, to the best of our knowledge.

[15] As for the Test Battery for Auslan Morphology and Syntax results for tests 4, 5, 9 and partially for test 6 have been published (Johnston, 2001; Schembri, 2001; Schembri et al., 2002). However, there was not sufficient funding to do all the work required (A. Schembri, personal communication, March 23, 2017).

The BSL Narrative Production Test also distinguishes between several domains (content, narrative structure, grammar), and several grammar features are assessed (see section 6.2, Research on SLI, for a discussion on researches that use this tool to investigate children with SLI).

The ASLAI has also several tasks to assess different linguistic features. The evolution of this tool is interesting because it contributes to the understanding of the difficulties that production assessment entails. The version described by Haug (2008) contained several tasks that asked the participant to retell a story or to describe an image or event. However, in the last version described by Enns et al. (2016) no similar task is included. Instead of eliciting narratives from participants, expressive skills are assessed by means of sign-sentence matching. The reason for this change was that the expressive tasks were too difficult to code and score properly, because of the difficulties they encountered to determine correctness of grammatical structure (R. Hoffmeister, personal communication, March 22, 2017).

Undoubtedly, assessment tools that provide a measure of sign language proficiency (as BSL-RST or ASL-SRT) are of great importance, since they allow to identify poor-performance and therefore they may reflect language disorders. However, when trying to deepen on the nature of poor-performance they may fall short. This is why some researches chose to use several instruments to approach the phenomenon or new tools are designed.

6.2. Research on language disorders in deaf signers

In this section, we review studies conducted in the field of deaf signers with language disorders due to specific language impairment (SLI), aphasia, or neurodegenerative disease.

One of the two existent BSL Sentence Repetition Tests was designed specifically to assess the language ability of deaf children suspected to have SLI (C. R. Marshall, Mason, et al., 2015). Unlike the ASL-SRT, it controls for the presence of agreement verbs and classifiers to analyze the performance in these features (out of 17 items, 5 are classifier constructions and 5 contain agreement verbs). The authors argue that "Because previous studies of sentence repetition in spoken languages have highlighted verb morphology as a significant difficulty for children with SLI, we looked in more detail at this feature in BSL" (C. R. Marshall, Mason, et al., 2015). It is the first research on sentence repetition with deaf signer children with SLI. By contrast, this field has been widely investigated with hearing population with SLI, and with orally-communicating children with hearing impairment. Regarding verb agreement and classifiers, the authors observed that either they were omitted, as in example (4), or they were substituted, as in example (5), where the semantic classifier substituted by a descriptive classifier:

(4) *Target*: TEDDY CL-curved-hand-ON TOP OF CUPBOARD BOY WANT WANT
'there is a teddy-bear on top of the cupboard that the boy really wants'
Answer: CUPBOARD BOY WANT WANT WANT

(5) *Target*: BOOK LOTS CL-flat-hand-ROW ROW ROW
'lots of books in a line'
Answer: BOOK CL- tracing-handshape-LINE ROW

The authors also provide the following example of substitution in sentences with agreement verbs, since they consider that ALL-OVER-FLOOR substitutes POUR-ON-FLOOR:

(6) *Target*: BOY NAUGHTY BATH WATER CUP-SCOOP-WATER POUR-ON-FLOOR
'there is a naughty boy in a bath and he throws water on the floor'

Answer: NAUGHTY BOY, BATH, ALL-OVER-FLOOR

In this cross-sectional study, they observed that scores increased with age in the control group, but those of the SLI group did not. These results are in line with those of hearing SLI chil-

dren, and of children with hearing impairment who communicate orally and who were tested orally.

This study is an example of those that combine the use of several sign language assessment tools: the BSL-RST, the BSL production test and the nonsense sign repetition test.

Results from the BSL-RST reveal poor performance of SLI participants, but no detail is given about which structures are more challenging for them.

As discussed above, the fact that the test was designed to study the BSL development, not specifically to study or diagnose SLI, may lead to difficulties in deepening on interpretation of the data obtained with this tool. The BSL production test results provide some more data to understand the phenomenon of SLI in signers. Most of participants (n = 13) failed in narrative content, narrative structure and grammar elements studied in the test. The authors present an example of the beginning of a narrative produced by a deaf SLI child (Mason et al., 2010, p. 43):

(7) *Typically developing child (aged 13;11)*:
WHEN FIRST BOY LIE-DOWN-REST ON SOFA IN LIVING ROOM WATCH TV WATCH HIS SISTER 'SHRUG' GIRL BRING-TRAY PUT-DOWN FOOD THERE ORANGE JUICE (POINT TO LOCATION) PLATE CAKE (POINT TO LOCATION) PLATE BREAD CLASSIFIER (FLAT OBJECT-BREAD).

SLI child (aged 12;09):
SIT SIT BOY LAZY WATCH TV HE DEMAND DEMAND

The longitudinal study of a signer who was given the name of Adam conducted by Quinto-Pozos et al. (2016) is also of great interest. They conducted a case-study research during 4 years that involved interviews with the teenager's family, professionals working with him, a neuropsychological evaluation, and they assessed his ASL using the ASL-PA and the ASL-RST. Although these assessments were designed for younger children, they used them "because it is not uncommon for children with language impairment to perform like younger users of a language" (Quinto-Pozos et al., 2016, p. 4).

Results from ASL-PA administration reveals, however, that "one can also do well on the ASL-PA with short conversational turns that are grammatical and do not tax sequential memory" (Quinto-Pozos et al., 2016, p. 8). As for results from ASL-SRT, the subject (Adam) scored below native adults and children signers, comparing to the norms for native signers. Although a slight performance improvement was observed between the three elicitation periods, all assessments showed scores significantly below the 1.25 SD (the cut-off used in some studies for language impairment). In the qualitative analysis of results, the authors observe that over time Adam omits fewer target signs and alters fewer word order. Fingerspelling is challenging for him, and longer and more complex sentences are more difficult for him to reproduce.

Another case study, that of Paul, was reported by Morgan, Herman, & Woll (2007). Paul was a native deaf signer aged 5;2. He performed 1.3 standard deviations below the mean for the BSL-RST. He would meet the criteria followed by Quinto-Pozos et al. (2016), but Morgan et al. consider the cut-off as being 1.5-2 SD below the mean. He expressed well plurals (which in BSL is more lexical than morphological), but struggled with constructions as negation, noun-verb distinctions, spatial verbs, and classifiers. The authors point out that these structures are related to forms encoded through morphosyntactic rules (head-shake for negation, phonological changes in noun-verb distinctions, complex embedding of morphological information in spatial verbs and classifiers). In the BSL Narrative Production Test he was in the 25th centile for all skills (narrative content, narrative structure and BSL grammar). He produced single signs with few inflections, as in the following example, lacking verb agreement (compared to a typically developing native signer):

(8) *Target picture: a man giving a boy a letter*
Typically developing native signer aged 4;6

Child: MAN LETTER GIVE-3rd person
'the man gives the letter to (him/her)'

Paul: GIVE GIVE SQUARE GIVE (citation forms)
A: SQUARE GIVE WHO?
C: GIVE GIVE POINT (picture) LETTER
A: PICTURE WHAT?
C: LETTER POINT

The research conducted by Herman et al. (2014) is, according to the authors, the first study on narrative abilities of deaf signing children with SLI. The deaf signing children with SLI scored significantly lower than the control group (17 deaf children with BSL as their first language), in measures of content, narrative structure, and BSL grammar. The only grammatical feature that the authors mention as significantly different between the two groups is the use of classifiers. As described by Morgan, Herman, & Woll (2007), narratives produced by children with SLI tend to have less simultaneity (utterances are more linear/sequential) and contain less morphological inflections. Significant differences between narratives by both groups in narrative structure were: less information about the climax, the resolution of the story, and the character's motivations (evaluative comment) in the SLI group.

Atkinson et al. (2005) tested verb, locative sentence, and classifier comprehension in three sentence/verb phrase–picture matching tasks. The aim of the study was research on sign language comprehension laterality. Both groups, participants with left hemisphere (LH) and right hemisphere (RH) damage, were impaired for classifier and locative sentence comprehension. Differences between the groups emerged in the single verbs and sentences comprehension task, which supports the idea that sign language comprehension is, like spoken language, LH dominant (see also Hickok, Bellugi, & Klima, 1998; Hickok, Love-Geffen, & Klima, 2002).

Hicock et al. (2009) used a narrative production task to compare the production of ASL classifier constructions and lexical signs in participants with either left or right hemisphere damage. They showed that whereas LHD participants produced fewer signs overall, they produced fewer classifier errors than RHD patients. In contrast, RHD participants had a significant number of classifiers errors, with almost no lexical errors. However, the authors did not analyze the types of errors produced in the classifier constructions, so it is unclear whether the errors were in handshape (lexical) or in the motion of signs.

6.3. Research on first language late exposure

Research on this field has been conducted mainly using tasks designed specifically for these researches.

The sentence recall task used by Mayberry and colleagues (Mayberry, 1993; Mayberry & Eichen, 1991; Mayberry & Fischer, 1989) supported the hypothesis that first language late exposure affects sentence processing skills: lexical recall, word order recall, preservation of the target sentence meaning, and grammatical acceptability.

Boudreault and Mayberry (2006) deepened on the nature of the effects of first language late exposure. They compared grammaticality judgement accuracy. Stimuli were examples of 6 syntactic structures: simple sentences, negative sentences, sentences containing agreement verbs, wh-questions, relative clauses, and classifier sentences. They found that the most accurately judged structure was negative sentences, and the structure that was most inaccurately judged was relative clauses. Results do not support one of the hypotheses of their project, that is, first structures acquired by children (simple, negative, and agreement verb structures) would be those presenting more accurate results, whereas structures acquired in later stages (wh-questions, relative clauses, and classifier predicates) would be judged less accurately.

Results from Cormier et al. research (2012) are in line with the previous study. Moreover, they found that "the ungrammatical agreeing sentences appear to be relatively 'easy' to identify" (Cormier et al., 2012, p. 62), whereas ungrammaticality in relative clauses was less accurately detected.

7. Final remarks

The review of previous research on language assessment of deaf signers with SLI or aphasia and of late learners of sign language as a first language, leads us to consider the following issues in the design of the SIGN-HUB tools.

As a first point, although various tools to assess sign language skills have been built, no battery of tests exists that covers all main aspects of a sign language (phonological, morphological, semantic, syntactic, pragmatic, narrative knowledge). Tests designed so far focus on specific aspects of linguistic competence by leaving other out of the picture. This reflects the fact that many aspects of sign language grammar are still severely underinvestigated and it is hard to build assessment tests in absence of this knowledge. This observation holds for SIGN-HUB task 2.3, although we explicitly mitigate this problem by working in close collaboration with researchers in task 2.1, who will provide grammar descriptions which are necessary to build the relevant assessment tools. Having said that, given time and human resources limitations, we cannot build a comprehensive assessment tools either, so choices are required, as planned from the beginning of the project. In particular, we focus on *some* lexical tests and on *some* syntactic tests.

Let us start from lexical tests. Our starting point was the British Sign Language Vocabulary, the most advanced tool in this area.

There are aspects of the British Sign Language Vocabulary Test that researchers in SIGN-HUB task 2.3 might want to capitalize on:

(i) choice of core lexicon items
(ii) strategies to minimize the effect of iconicity in picture-matching task
(iii) use of phonological and semantic distractors
(iv) production tasks based on meaning association.

On the other hand, some aspects of the tests might be reconsidered.

(i) While the use of the very same target items in all the four tasks allowed a comparison of their difficulty, it introduces the potential problem that the same item can be recalled from a previous task, especially if more than one task is administered in the same session.
(ii) The presence of a phonological and a semantic distractor allows to deeply probe the role of phonological and semantic distractors (for example by studying if the phonological similarity in a given parameter is more disruptive than phonological similarity in a different parameter). They also allow to identify the nature of the lexical difficulty that a person has - and track its origin. However, building two distinct comprehension tasks, one with phonological distractors and a second one with semantic distractors, might allow to do this better than having the phonological and semantic distractors in the same test.
(iii) It is conceivable that the production task based on meaning association may give ambiguous results ("bone" is associated to "dog", but can "garden" be considered an associate?). A more focused meaning association (say, "produce the opposite") might be less ambiguous.

Regarding syntactic tests, a review of the existent tools shows that research in this area is even more fragmentary and suggests that it is a safe strategy to restrict our attention to a limited number of « litmus constructions », namely to linguistic aspects that are known to be sensitive to differences in populations and language skills. In order to choose these litmus constructions the following aspects need to be taken into consideration:

(i) Linguistic features to be assessed: a general goal is disentangling poor-performance due to delayed first language exposure from poor-performance due

to other language disorders (SLI, aphasia, etc.). In order to do this a limited set of constructions should be selected that are complex enough to distinguish between these types of populations. In principle, one way to identify them is studying how different constructions develop, at what age and in what order. However, very little is known about the order of acquisition of syntactic constructions in sign languages and studying it is simply beyond the scope of the project. So, another strategy seems more viable. On the one hand, we will build the test involving those constructions that are known to be problematic for special populations cross-linguistically *in spoken languages*. There is a large body of literature showing that so called A'-dependencies (especially object questions and object relatives) are particularly challenging for special populations and may reveal relatively subtle impairments (Friedmann, Belletti, & Rizzi, 2009; Friedmann & Novogrodsky, 2011; among others). As A'-dependencies are relatively well studied in the sign languages for which we are going to build assessment tools, we will be working under the assumption that they are a good example of litmus constructions in sign languages, although we are ready to revise this assumption if our research shows otherwise.

As focusing only on A'-dependencies might leave out of the picture constructions that are sign language specific (or that are more prominent in sign languages than in spoken languages) we plan to build tests to study other constructions that are at least anecdotically complex, such as (spatial) agreement and role shift.

No specific test is planned for phonological or morphological competence. This is mostly due to a realistic evaluation of the resources that are available. However, we would like to stress that the lexical tasks do not only grasp lexical skills in general, but also fine grained phonological competence and conceptual semantic competence. As for the syntactic tests, the agreement and role shift tasks involve morphological competence in addition to plain syntactic ones.

(ii) To make sure that tests will be sensitive enough to difficulties in specific linguistic features, enough items should be designed for each feature.

(iii) Time spent in scoring production skills: to assess both receptive and expressive skills is advisable, but scoring the latter may be very time consuming. The ASLAI option can help overcome these disadvantages.

(iv) Assessment tools and discourse: assessing production skills by means of retelling tasks may be too time consuming. However, elicitation of a short narrative may be advisable. Tests based on closed-ended questions focus on lexical and sentence levels. Counting with a narrative by the test taker, although it is not analyzed thoroughly, would allow researchers to contextualize quantitative, lexical and sentence level, data.

As for participants' recruitment, the following aspects should be considered:

(i) As far as possible, we should ensure that participants are "true cases of delayed L1 acquisition of a sign language, as opposed to second language (L2) acquisition of a sign language after successful acquisition of the surrounding spoken/written language as L1" (Cormier et al., 2012, p. 51). We will carefully collect background information about the early use of spoken and signed languages for each participant to distinguish between sign language as L1 or L2, and to examine the effect of the age of L1 exposure.

(ii) Questionnaires to be filled by teachers or speech therapists may be useful for the sampling process (Mason et al., 2010).

8. References

Alba de la Torre, C. (2016). *Wh-questions in Catalan sign language*. Universitat Pompeu Fabra, Barcelona. Retrieved from http://www.tesisenred.net/handle/10803/397751

Anderson, D., & Reilly, J. (2002). The MacArthur Communicative Development Inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education*, *7*(2), 83–106.

Atkinson, J., Marshall, J., Woll, B., & Thacker, A. (2005). Testing comprehension abilities in users of British Sign Language following CVA. *Brain and Language*, *94*(2), 233–248.

Bel, A., Ortells, M., & Morgan, G. (2014). Reference control in the narratives of adult sign language learners. *The International Journal of Bilingualism*, 1–17.

Bellugi, U., Klima, E. S., & Siple, P. (1975). Remembering in signs. *Cognition*, *3*(2), 93–125.

Bizer, S., & Karl, A.-K. (2002). *Entwicklung eines Wortschatztests für gehörlose Kinder im Grundschulalter in Gebärden-, Schrift-und Lautsprache*. Hamburg.

Boudreault, P., & Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes*, *21*(5), 608–635.

Cormier, K., Schembri, A., Vinson, D., & Orfanidou, E. (2012). First language acquisition differs from second language acquisition in prelingually deaf signers: Evidence from sensitivity to grammaticality judgement in British Sign Language. *Cognition*, *124*(1), 50–65.

Cormier, K., Smith, S., & Zwets, M. (2013). Framing constructed action in British Sign Language narratives. *Journal of Pragmatics*, *55*, 119–139.

Denmark, T., & Atkinson, J. (2015). Using and developing language and cognitive assessments with deaf signers. In E. Orfanidou, B. Woll, & G. Morgan (Eds.), *Research Methods in Sign Language Studies. A Practical Guide* (pp. 352–368). Malden, MA: Wiley-Blackwell.

Enns, C., Haug, T., Herman, R., Hoffmeister, R., Mann, W., & McQuarrie, L. (2016). *Exploring Signed Language Assessment Tools in Europe and North America*. *Diversity in Deaf Education* (Vol. 6).

Enns, C., & Herman, R. (2011). Adapting the Assessing British Sign Language Development: Receptive Skills Test into American sign language. *Journal of Deaf Studies and Deaf Education*, *16*(3), 362–374.

Frederiksen, A. T., & Mayberry, R. I. (2014). Who's on First? Investigating the referential hierarchy in simple native ASL narratives. *Lingua*, *180*, 49–68.

Frederiksen, A. T., & Mayberry, R. I. (2015). Tracking Reference in Space: How L2 Learners Use ASL Referring Expressions. In *Proceedings of the 39th annual Boston University Conference on Language Development.* (pp. 165–177). Cascadilla Press.

Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, *119*(1), 67–88.

Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand? The comprehension of Wh questions in three subtypes of SLI. *Lingua*, *121*(3), 367–382.

Geraci, C., Gozzi, M., Papagno, C., & Cecchetto, C. (2008). How grammar can cope with limited short-term memory: Simultaneity and seriality in sign languages. *Cognition*, *106*(2), 780–804.

Givón, T. (1984). The pragmatics of referentiality. In D. Schiffrin (Ed.), *Georgetown University Round Table on Languages and Linguistics 1984* (pp. 120–138). Washington, D.C.: Georgetown University Press.

Göksel, A., & Pfau, R. (n.d.). The non-native lexicon. In J. Quer, C. Cecchetto, C. Donati, C. Geraci, M. Kelepir, R. Pfau, & M. Steinbach (Eds.), *SignGram Blueprint: A guide to sign language grammar writing*. Mouton de Gruyter.

Goldstein, G., & Bebko, J. M. (2003). The Profile of Multiple Language Proficiencies: A Measure for Evaluating Language Samples of Deaf Children. *Journal of Deaf Studies and Deaf Education*, *8*(4), 452–463.

Haug, T. (n.d.). Sign Language Assessment. Retrieved March 19, 2017, from http://www.signlang-assessment.info/

Haug, T. (2008). Review of Sign Language Assessment Instruments. In A. Baker & B. Woll (Eds.), *Sign Language Acquisition* (pp. 51–86). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Haug, T. (2011a). *Adaptation and Evaluation of a German Sign Language Test. A Computer-Based Receptive Skills Test for Deaf Children Ages 4-8 Years Old*. Hamburg: Hamburg University Press.

Haug, T. (2011b). Methodological and theoretical issues in the adaptation of sign language tests: An example from the adaptation of a test to German Sign Language. *Language Testing*.

Hauser, P. C., Paludneviciene, R., Riddle, W., Kurz, K. B., Emmorey, K., & Contreras, J. (2016). American Sign Language Comprehension Test: A Tool for Sign Language Researchers. *Journal of Deaf Studies and Deaf Education*, *21*(1), 64–69.

Hauser, P. C., Paludneviciene, R., Supalla, T., & Bavelier, D. (2008). American Sign Language – Sentence Reproduction Test: Development & Implications. In R. M. 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* (pp. 155–167). Petrópolis, Brazil: Editora Arara Azul.

Hauser, P. C., Quinto-Pozos, D., & Singleton, J. L. (2015). Studying Sign Language Disorders: Considering Neuropsychological Data. In E. Orfanidou, B. Woll, & G. Morgan (Eds.), *Research Methods in Sign Language Studies: A Practical Guide* (pp. 336–351). Malden, MA: Wiley-Blackwell.

Herman, R. (2002). *Assessing British Sign Language Development*. City University.

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing BSL development: Receptive skills test*. Forest Books, Coleford.

Herman, R., Rowley, K., Mason, K., & Morgan, G. (2014). Deficits in narrative abilities in child British Sign Language users with specific language impairment. *International Journal of Language and Communication Disorders*, *49*(3), 343–353.

Hermans, D., Knoors, H., & Verhoeven, L. (2010). Assessment of sign language development: the case of deaf children in the Netherlands. *Journal of Deaf Studies and Deaf Education*, *15*(2), 107–119.

Hickok, G., Bellugi, U., & Klima, E. S. (1998). The neural organizarion of language: evidence from sign language aphasia. *Trends in Cognitive Sciences*, *2*(4), 129–136.

Hickok, G., Love-Geffen, T., & Klima, E. S. (2002). Role of the left hemisphere in sign language comprehension. *Brain and Language*, *82*(2), 167–178.

Hickok, G., Pickell, H., Klima, E. S., & Bellugi, U. (2009). Neural Dissociation in the Production of Lexical versus Classifier Signs in ASL: Distinct Patterns of Hemispheric Asymmetry. *Neuropsychologia*, *47*(2), 382–387.

Hoffmeister, R. J., Fish, S., Henner, J., Rosenburg, P., Benedict, R., & Caldwell-Harris, C. (2015). ASLAI. Retrieved March 21, 2017, from http://www.asleducation.org/pages/aslai.html

Hosemann, J. (2015). *The processing of German Sign Language sentences. Three event-related potential studies on phonological, morpho-syntactic, and semantic aspects*. Georg-

August-Universität, Göttingen.

Istituto di Scienze e Tecnologie della Cognizione - Consiglio Nazionale delle Ricerche. (n.d.). La piattaforma VOLIS. Guida per gli operatori. Retrieved March 22, 2017, from https://www.volis.it/index.php/guida-operator1

Jansma, S., Baker, A. E., & Knoors, H. (1997). Sign Language Assessment: A Dutch Project. *Deafness and Education*, *21*, 39–46.

Johnston, T. (2001). Nouns and Verbs in Austalian Sign Language: An Open and Shut Case? *Journal of Deaf Studies and Deaf Education*, *6*(4), 235–257.

Johnston, T. (2004). The assessment and achievement of proficiency in a native sign language within a sign bilingual program: The pilot Auslan receptive skills test. *Deafness & Education International*, *6*(2), 57–81.

Jones, A. C., Toscano, E., Botting, N., Marshall, C. R., Atkinson, J. R., Denmark, T., … Morgan, G. (2016). Narrative skills in deaf children who use spoken English: Dissociations between macro and microstructural devices. *Research in Developmental Disabilities*, *59*, 268–282.

Klatter-Folmer, J., van Hout, R., Kolen, E., & Verhoeven, L. (2006). Language development in deaf children's interactions with deaf and hearing adults: A dutch longitudinal study. *Journal of Deaf Studies and Deaf Education*, *11*(2), 238–251. https://doi.org/10.1093/deafed/enj032

Kotowicz, J., Turska, M. K., Woll, B., Herman, R., Łacheta, J., & Schromová, M. (2016). The adaptation of British Sign Language (BSL) Receptive Skills Test into Polish Sign Language (PJM), (December), 27–28.

Lichtig, I., Couto, M. I. V, Mecca, F., Hartley, S., Wirz, S., & Woll, B. (2011). Assessing deaf and hearing children's communication in Brazil. *J. Commun. Disord.*, *44*(2), 223–235.

Lieberman, A. M., & Mayberry, R. I. (2015). Studying Sign Language Acquisition. In E. Orfanidou, B. Woll, & G. Morgan (Eds.), *Research Methods in Sign Language Studies: A Practical Guide* (pp. 281–299). Oxford: John Wiley & Sons.

Maller, S., Singleton, J. L., Supalla, S., & Wix, T. (1999). The development and psychometric properties of the American sign language proficiency assessment (ASL-PA). *Journal of Deaf Studies and Deaf Education*, *4*(4), 249–269.

Mann, W. (2006). *Examining German Deaf Children's understanding of referential distinction in written German and German Sign Language (DGS)*. Tesi doctoral. University of California, Berkeley.

Mann, W. (2007). German deaf children's understanding of referential distinction in written German and German Sign Language. *Educational & Child Psychology*, *24*(4), 59–76.

Mann, W., & Haug, T. (2014). Mapping out guidelines for the development and use of sign language assessment: Some critical issues, comments and suggestions. In D. Quinto-Pozos (Ed.), *Multilingual Aspects of Signed Language Communication and Disorder* (pp. 123–142). Multilingual Matters.

Mann, W., & Marshall, C. R. (2012). Investigating deaf children's vocabulary knowledge in British Sign Language. *Language Learning*, *62*(4), 1024–1051.

Mann, W., Marshall, C. R., Mason, K., & Morgan, G. (2010). The Acquisition of Sign Language: The Impact of Phonetic Complexity on Phonology. *Language Learning and Development*, *6*(1), 60–86.

Marschark, M., Mouradian, V., & Halas, M. (1994). Discourse rules in the language productions of deaf and hearing children. *Journal of Experimental Child Psychology*, *57*(1), 89–107.

Marshall, C. R., Jones, A., Denmark, T., Mason, K., Atkinson, J., & Morgan, G. (2015). Deaf children's non-verbal working memory is impacted by their language experience. *Frontiers in Psychology*, *6*(May), 1–12.

Marshall, C. R., Mason, K., Rowley, K., Herman, R., Atkinson, J., Woll, B., & Morgan, G. (2015). Sentence repetition in deaf children with specific language impairment in British Sign Language. *Language Learning and Development*, *11*(3), 237–251.

Marshall, C. R., & Morgan, G. (2015). Investigating sign language development, delay and disorder in deaf children. In *The Oxford Handbook of Deaf Studies in Language* (p. 311). Oxford University Press.

Marshall, C. R., Rowley, K., & Atkinson, J. (2014). Modality-dependent and -independent factors in the organisation of the signed language lexicon: insights from semantic and phonological fluency tasks in BSL. *Journal of Psycholinguistic Research*, *43*(5), 587–610.

Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., & Woll, B. (2004). Aphasia in a user of British Sign Language: Dissociation between sign and gesture. *Cognitive Neuropsychology*, *21*(December 2014), 537–554.

Mason, K., Rowley, K., Marshall, C. R., Atkinson, J. R., Herman, R., Woll, B., & Morgan, G. (2010). Identifying specific language impairment in deaf children acquiring British Sign Language: Implications for theory and practice. *The British Journal of Developmental Psychology*, *28*(1), 33–49.

Mayberry, R. I. (1993). First-Language Acquisition After Childhood Differs From Second-Language Acquisition: The Case of American Sign Language. *Journal of Speech, Language, and Hearing Research*, *36*(6), 1258–1270.

Mayberry, R. I., & Eichen, E. B. (1991). The Long-Lasting Advantage of Learning Sign Language in Childhood: Another Look at the Critical Period for Language Acquistion. *Journal of Memory and Language*, *30*(4), 486–512.

Mayberry, R. I., & Fischer, S. D. (1989). Looking through phonological shape to lexical meaning: the bottleneck of non-native sign language processing. *Memory & Cognition*, *17*(6), 740–754.

Mayberry, R. I., & Witcher, P. (2005). What Age of Acquisition Effects Reveal about the Nature of Phonological Processing. *CRL Technical Reports*, *17*(3), 3–9.

McQuarrie, L., & Abbott, M. (2013). Bilingual Deaf Students' Phonological Awareness in ASL and Reading Skills in English. *Sign Language Studies*, *14*(1), 80–100.

Morgan, G. (1998). *The development of discourse cohesion in British sign language*. Tesi doctoral. University of Bristol.

Morgan, G., Herman, R., & Woll, B. (2007). Language impairments in sign language: breakthroughs and puzzles. *International Journal of Language & Communication Disorders*, *42*(1), 97–105.

Niederberger, N. A. (2004). *Capacités langagières en langue de signes française et en français écrit chez l'enfant sourd bilingue: quelles relations?* Tesi doctoral. Université de Genève.

Quinto-Pozos, D., Singleton, J. L., & Hauser, P. C. (2016). A Case of Specific Language Impairment in a Deaf Signer of American Sign Language. *Journal of Deaf Studies and Deaf Education*, 1–15.

Reynolds, W. (2016). *Early bimodal bilingual development of ASL narrative referent cohesion: Using a heritage language framework*. Gallaudet University.

Sánchez Amat, J. (2015). *Llengua de signes i llengua escrita en la modalitat educativa bilingüe i en la intervenció amb l'infant sord*. Universitat Autònoma de Barcelona, Bellaterra. Retrieved from https://www.academia.edu/21990092/Llengua_de_signes_i_llengua_escrita_en_la_modalitat_educativa_bilingüe_i_en_la_intervenció_amb_linfant_sord

Schembri, A. (2001). *Issues in the analysis of polycomponential verbs in Australian Sign Language (Auslan)*. University of Sydney, Sydney.

Schembri, A., Wigglesworth, G., Johnston, T., Leigh, G., Adam, R., & Barker, R. (2002). Issues in

Development of the Test Battery for Australian Sign Language Morphology and Syntax. *Journal of Deaf Studies and Deaf Education*, *7*(1), 18–40.

SignMET. (2013). *SignMET Final Report*. Retrieved from http://www.signmet.eu/SignMET/pdf/5_3-SignMET-Final-Report.pdf

Simms, L., Baker, S., & Clark, M. D. (2013). The Standardized Visual Communication and Sign Language Checklist for Signing Children. *Sign Language Studies*, *14*(1), 14–17.

Singleton, J. L., & Supalla, S. J. (2003). Assessing Children's Proficiency in Natural Signed Languages. *The Oxford Handbook of Deaf Studies, Language, and Education*, *1*, 289–302.

Stein, N. L., & Glenn, C. G. (1975). An Analysis of Story Comprehension in Elementary School Children: A Test of a Schema. Retrieved from http://eric.ed.gov/?id=ED121474

Strong, M., & Prinz, P. (1997). A study of the relationship between American Sign Language and English literacy. *Journal of Deaf Studies and Deaf Education*, *2*(1), 37–46.

Supalla, T., Hauser, P. C., & Bavelier, D. (2014). Reproducing American Sign Language sentences: Cognitive scaffolding in working memory. *Frontiers in Psychology*, *5*(AUG), 1–16.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, *24*(5), 595–611.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612–630.

Valmaseda, M., Pérez, M., Herman, R., & Ramírez, N. (2013). Evaluación de la competencia gramatical en LSE: Proceso de adaptación del BSL-Receptive Skill Test (test de habilidades receptivas). In *Congreso CNLSE sobre la investigación de la lengua de signos española*. Madrid: CNLSE. Retrieved from http://www.cnlse.es/es/virtual-library/evaluación-de-la-competencia-gramatical-en-lse-proceso-de-adaptación-del-bsl

Vercaingne-Menard, A., Godard, L., & Labelle, M. (2001). The emergence of narrative discourse in two young deaf children. In V. Dively, M. A. Metzger, S. Taub, & A. M. Baer (Eds.), *Signed languages: Discoveries from international research* (pp. 120–134). Gallaudet University Press.

Wilson, M., & Emmorey, K. (1997). Working memory for sign language: A window into the architecture of the working memory system. *Journal of Deaf Studies and Deaf Education*, *2*(3), 121–130. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15579841

Wilson, M., & Emmorey, K. (1998). A "word length effect" for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition*, *26*(3), 584–590.

Woolfe, T., Herman, R., Roy, P., & Woll, B. (2010). Early vocabulary development in deaf native signers: A british sign language adaptation of the communicative development inventories. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *51*(3), 322–331.

Zwitserlood, I. (2012). Classifiers. In *Sign language: An international handbook* (pp. 158–186). Mouton de Gruyter.