

The Corpus of Israeli Sign Language

Rose Stamp , Ora Ohanin & Sara Lanesman

Bar-Ilan University
Department of English Literature & Linguistics
rose.stamp@biu.ac.il

Abstract

The Corpus of Israeli Sign Language is a four-year project (2020-2024) which aims to create a digital open-access corpus of spontaneous and elicited data from a representative sample of the Israeli deaf community. In this paper, the methodology for building the Corpus of Israeli Sign Language is described. Israeli Sign Language (ISL) is the main sign language used across Israel by around 10,000 people. As part of the corpus, data will be collected from 120 deaf ISL signers across four sites in Israel: Tel Aviv and the Centre, Haifa and the North, Be'er Sheva and the South and Jerusalem and the surrounding area. Participants will engage in a variety of tasks, eliciting a range of signing styles from free conversation to lexical elicitation. The dataset will consist of recordings of over 360 hours of video data which will be used to conduct sociolinguistic investigations of language contact, variation, and change in the near term, and other linguistic analyses in the future.

Keywords: Israeli Sign Language, corpus, lexical database, corpus project, language contact, language change

1. Introduction

Corpora provide spontaneous, naturalistic data against which claims about the structure and use of a given language can be tested. The need for sign language corpora is of paramount importance because they remain under researched compared to spoken languages. Furthermore, it is important to widen the set of languages represented in corpus linguistics (McEnery & Ostler, 2000). While there has been a surge in sign language corpora creation since the turn of the century, with the addition of over twenty around the globe (Kopf et al., 2021), there are still many sign languages without existing corpora, especially those located in the Middle East. To add to the diversity of sign language corpora and to better understand sign language variation in Israel more specifically, the Corpus of Israeli Sign Language project was launched in 2020.

The Corpus of Israeli Sign Language is a four-year project (2020-2024) funded by the Israeli Science Foundation and hosted by Bar-Ilan University. The primary objective of the project is to conduct sociolinguistic studies on language contact, variation, and change in Israeli Sign Language (ISL), as described in Section 2. To achieve this, a machine-readable digital corpus of spontaneous and elicited data from the Israeli deaf community will be created. The project is led by Dr. Rose Stamp, together with her research team, Ora Ohanin and Sara Lanesman, who are native signers of ISL. In this paper (in Section 3), we outline the methodologies for collecting a representative sample of language data from the ISL deaf community, including the sampling method, stimuli, and task procedures. The methodology follows other sign language corpora around the world, drawing on a combination of tasks used in the British Sign Language (BSL) Corpus Project, the German Sign Language (DGS) Corpus Project and others.

Finally, in Section 4, we describe the related resources, including the project's website, the lexical database hosted by Global SignBank, and the online network-based visualization website, ISL-LEX. The Corpus of ISL will provide one of the first large-scale datasets of a young sign language and will serve as a key resource for researchers investigating ISL structure and usage.

2. Israeli Sign Language & the Deaf Community

Compared to other countries, Israel has a unique abundance of sign languages, which emerged naturally within the last hundred years (Meir & Sandler, 2008). The main sign language used in Israel is Israeli Sign Language (ISL), with an estimated 10,000 users. It is the language of the National Deaf Association, the education system, and sign language interpreting. ISL is a relatively young sign language, roughly about 90 years old, which arose with the formation of the deaf community in Israel around the 1930s, beginning with the establishment of the first school for the deaf in 1932 in Jerusalem. Many of the first generation ISL signers, who are now the older population in Israel, immigrated from Europe, North Africa and the Asia and were illiterate or semi-literate when ISL first emerged. Today, in contrast, younger deaf people, who are the third or fourth generations of ISL signers, are multilingual and are exposed to a variety of signed, spoken and written languages. The deaf ISL community has undergone rapid changes due to increased mobility, exposure to different languages within the education system, and changes in social communication and technology. These changes have led to increased contact between languages and language varieties. The creation of the Corpus of ISL is an opportunity to capture the linguistic variation and to find clues to the social demographic forces involved.

The aim of this project entitled “A corpus-based sociolinguistic study of sign languages in Israel” is not just to create a corpus of ISL, but also to address important research questions regarding language contact, variation, and change. The project presents a systematic investigation of language contact in three different situations: (1) contact between ISL regional varieties, (2) contact between ISL and Arabic, and (3) contact between ISL and a local sign language used in Kufr Qassem.

In the first contact situation, we look at how increased mobility in recent years might have had an influence on sign language varieties across Israel. It is claimed that regionally distinct varieties exist in ISL; for example, there are at least two variants for the sign ‘hospital’, one

associated with signers from Tel Aviv, and one associated with signers from Haifa (see Figure 1).



Figure 1: Two variants for the sign ‘hospital’: (left) in Tel Aviv and (right) in Haifa

Studies show that increases in mobility and changes in communication patterns might result in a decline in the use of such regionally distinct signs (McKee & McKee, 2011; Stamp et al., 2014). This has been shown to lead to long-term language change (Trudgill, 1986), in particular dialect levelling, in which signers reduce their use of regionally marked variants in favour of variants that are distributed over a wider geographical area (Kerswill, 2003). In this first study, we explore whether there is any evidence of language change in ISL because of increased contact between language varieties.

In the second study, we look at the unique contact situation between ISL and Arabic. For many years, deaf children from Arabic-speaking families were educated in the Jewish sector, and they were exposed to ISL and Hebrew. Nowadays, deaf children from Arabic-speaking families either attend mainstream schools alongside their hearing peers or schools in which ISL and Arabic are the main languages of instruction. While the effects of contact between ISL and Hebrew have been described previously (Meir & Sandler, 2008), few studies have focused on the contact situation between ISL and Arabic. In addition, deaf children in Arab communities are exposed to Arabic in its diglossic form. In other words, children are exposed to two distinct forms: one colloquial spoken variety and one formal written variety (Saiegh-Haddad, 2012). We investigate the contact situation between ISL and the two distinct varieties of Arabic and how each variety might affect the other.

The third study investigates the language contact situation in Kufr Qassem, an Arabic-speaking community situated in the Southern Triangle area in Central Israel, around 20 kilometers northeast of Tel-Aviv. Younger deaf people in Kufr Qassem are exposed to two sign languages: ISL, as the dominant sign language used across Israel and Kufr Qassem Sign Language (KQSL), a local sign language used only by the deaf community in Kufr Qassem (Kastner et al., 2014). KQSL emerged around 90 years ago when a high number of deaf people were born into the local community. First generation signers were relatively isolated and unaffected by other sign languages, as many did not attend school or live outside of the local community. However, the situation for third generation signers is very different. Because of increased mobility as well as changes to the instruction language used in the school for deaf children in Kufr Qassem, deaf people are now exposed to many languages, including KQSL, ISL, Arabic, Hebrew, etc. This has led to drastic language shift within this community. Recent findings suggest that younger deaf

signers in Kufr Qassem are dominant in ISL, and that the continuation of KQSL is under threat from language shift (Stamp & Jaraisy, 2021). The aim of this third study is to investigate the influence of ISL on smaller sign language communities across Israel, using Kufr Qassem as our case study.

To conduct sociolinguistic investigations, such as these, on language contact, variation, and change in ISL, it is crucial to analyze language data from a representative sample of the Israeli deaf community. The corpus will serve as a large-scale dataset of ISL, which will be digitalized, annotated and made publicly available for research as well as applied purposes.

3. Methodology

At the time of writing this paper, data collection is currently underway and therefore, in the following section, we describe the methods as planned, rather than based on what we have collected to date.

3.1 Participants

To collect a representative sample of the ISL deaf community, thirty deaf ISL signers will be recruited from four key sites across Israel. The sites represent the major areas in Israel, including Tel Aviv and the Centre, Haifa and the North, Be’er Sheva and the South and Jerusalem (the capital city) and the surrounding area. These sites were selected because they contain adequate numbers of deaf signers, and because they represent sites which vary in terms of their signing varieties (e.g., the sign for ‘hospital’ between Tel Aviv and Haifa). Deaf fluent signers of ISL were recruited; no criteria based on family background were required (native and non-native signers are included), however, metadata regarding this was collected from each individual, as described below.

Israel is a relatively small country and therefore many individuals spend time in multiple sites during their lifetimes. As a result, participants were filmed in the site in which they lived the most within the last ten years. In each site, ten participants were recruited in three age groups: younger (18-39), middle-aged (40-59), and older (60+). During the selection of participants, gender, social class, ethnicity (e.g., Jewish, Arab) and family origin (e.g., Moroccan, German) were considered, taking a representative sample when possible. See Table 1 for participant characteristics.

Region	Age groups			Gender	
	Young (18-39)	Middle (40-59)	Older (60+)	Male	Female
Haifa & the North	10	10	10	15	15
Tel Aviv & the Centre	10	10	10	15	15
Greater Jerusalem	10	10	10	15	15
Be’er Sheva & the South	10	10	10	15	15
Total	40	40	40	60	60

Table 1: Participant characteristics

Four deaf fieldworkers were recruited, one in each site. All fieldworkers are deaf native ISL signers who live in the respective target sites and have good contacts with the local deaf community. Their role was to identify and recruit ISL signers who live in the local community. In addition, a deaf fieldworker coordinator was recruited to oversee the data collection process and to serve as the consistent interviewer in one of the tasks (described in Section 3.2).

A technique of network sampling was used, in which fieldworkers began by recruiting people they know, then asked these people to recommend other individuals who matched the project criteria (Milroy & Gordon, 2003). In this way, participants were filmed in pairs consisting of two individuals who know each other.

A Deaf Advisory Committee was set up, consisting of six deaf ISL signers from different backgrounds, who are active and prominent figures in the Israeli deaf community. The committee serves as a consultation board for various issues related to variation in ISL, stimuli selection, website design, etc. In addition, Professor Adam Schembri, who was the PI for the BSL Corpus Project and consultant on several other corpora, is the International Consultant on this project, and he is providing us with advice on methodological issues, lemmatization, and glossing at various stages in the project.

3.2 Stimuli & Procedure

The Corpus of ISL follows the methodologies outlined in other sign language corpora, and, in particular, those from the BSL Corpus Project and the DGS Corpus Project.

The data were collected in two stages: first, an online meeting between each participant and the fieldworker, and then a 3-hour sociolinguistic interview, conducted in pairs onsite. Based on Labov's classic sociolinguistic interview (1972), we included seven language tasks which elicit a range of signing styles from spontaneous to elicited, including:

Stage 1 (online, one-to-one meeting):

1. short questionnaire about name signs,
2. lexical elicitation task,

Stage 2 (onsite, filmed in pairs):

1. personal narrative,
2. free conversation,
3. retelling of the events shown in a video clip,
4. questions and answers about language variation and change,
5. retelling of the events shown in short video clips,

As part of stage one, each participant met with the fieldworker online to complete two of the seven language tasks. In most cases, this was completed using the online platform, Google Meet. An online format was preferred because the commencement of filming coincided with the COVID pandemic when filming in person was not possible. During this meeting, participants first completed a consent form and video sharing consent form to agree that their data can be made openly accessible. Then, participants

completed a 43-item questionnaire about their language background, education, language preferences, etc. The items in the questionnaire were largely based on the questionnaire used in the BSL Corpus Project; however, it was adapted for the purposes of the Israeli deaf community and translated into Hebrew and Arabic. The questionnaires were completed using Google Forms. Following this, participants were asked about their name sign (name signs refer to the visual name given to members of the deaf community). Three questions were asked: (1) what is your name sign in ISL? (2) what is the reason for your name sign? (3) has your name sign stayed the same throughout your lifetime? As part of this project, we plan to conduct a diachronic analysis of name signs, similar to other studies (e.g., Börstell, 2017; McKee & McKee, 1999).

Finally, participants were asked to give their sign variants for a list of concepts. The aim of the lexical elicitation task was to elicit participant's preferred variants for the concepts known to vary considerably and to investigate how this variation patterns across different social groups. Participants were encouraged to give their preferred variant(s) and to mention other variants they know or have seen. For this reason, individual meetings were preferred to avoid the influence of one participant's answers upon another's. The concepts on the list were selected because they are known to show considerable variation in ISL and because the variation is claimed to be associated with social factors, such as a signer's age, gender, ethnicity, religion, regional background, and so on. The list was compiled with the help of the online ISL dictionary¹, which includes multiple variations for the same concept, and consultation with the Deaf Advisory Committee. In the end, the lexical elicitation task consisted of two slides acting as trials, followed by 145 slides for the actual task elicitation. Each slide showed a picture together with the sign's closest Hebrew and Arabic equivalent translations representing each of the target concepts (e.g., a coloured orange square with the Hebrew word כחום and Arabic برتقالي to elicit the sign for 'orange'). See Figure 2 below.

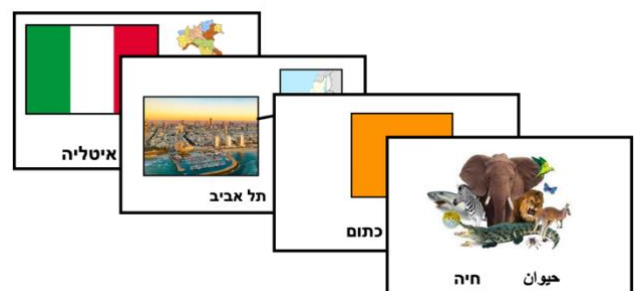


Figure 2: Example of the slides used in the Lexical Elicitation Task

In the second stage, a sociolinguistic interview was carried out. This consisted of the five remaining tasks, filmed at each site over a duration of three hours. Participants were filmed in pairs with another signer from the same age group and region. Filming took place in the local deaf club and on some occasions, in a classroom at Bar-Ilan University (for participants from the Tel Aviv site). The PI, the fieldworker

¹The ISL dictionary was originally created by The Institute for the Advancement of Deaf Persons in Israel (IADPI) and is now

hosted by Maggalei Shema (https://isl.danfishgold.com/#he-3SI).

coordinator, and the fieldworker themselves were present at all filming sessions. For the personal narrative task, participants were informed in advance that they should think of a personal narrative lasting 5 minutes. Each participant was filmed telling their narrative to their interlocutor. After 5 minutes, participants were informed, in a non-intrusive way, that their time had run out (but that they were able to conclude their story briefly). Then, participants switched, and the other participant told their narrative. Following this, both participants engaged in free conversation (Task 2). Participants were left in front of the cameras with no intervention for thirty minutes. The aim of this task is to elicit data as naturalistic as possible (Labov, 1972).

In the third task, participants took turns watching an excerpt from a movie clip and they were asked to retell the events to their interlocutor. One participant watched an excerpt from Charlie Chaplin's *The Lion's Cage* (*The Circus*, 1928) and the other watched an excerpt from an animated cartoon called *Snack Attack* (2012). Both movies were selected because they contain no dialogue and, while they are different storylines, they both elicit a range of sign language features including constructed action, character reference, and use of classifiers. *The Lion's Cage*, specifically, has also been used for elicitation in previous ISL projects and therefore it has been shown to be a successful form of elicitation and the data elicited in the corpus can be compared to previous elicitations (e.g., Stamp et al., 2018). To avoid issues related to memory, participants were first shown the whole clip from start to finish and then they were shown the clip in parts and asked to retell the events directly after each section they watched. After all the parts were retold and participants had fully internalized the storyline, they were asked to retell the whole story from start to finish. Participants were also informed that their partners would complete a comprehension task after their retelling, which involved ordering five movie stills in chronological order of the events as they were described. Awareness of the comprehension task encouraged participants to be more detailed in their retellings.

In task four, participants were interviewed about their patterns of language use, their attitudes towards different language varieties, and about their own examples of language variation and change in ISL. The fieldworker coordinator served as the consistent interviewer in each filming site. The interview was conducted with both participants simultaneously. Questions were modified from the BSL Corpus Project interview after consultation with their team. They included questions such as: Do you think there are differences in signing between older and younger ISL signers? If you moved to a new location in Israel, would change your signing to accommodate to those in the new location? The interview lasted 20 minutes.

In the final task, participants watched three excerpts from Sylvester and Tweety's *Canary Row* (1950) cartoon (a total of six altogether). Each excerpt ranged in duration from 18-31 seconds. After watching each excerpt, they retold the events to their interlocutor. As part of a comprehension task, the interlocutor was given three pictures, each representing one of the excerpts, and was asked to identify the picture which most resembled what they understood from the retelling. *Canary Row* was selected specifically

because the data is comparable with other sign language corpus projects including those in Germany, Spain, Poland and the Netherlands (Kopf et al., 2021). At the end of filming, all participants were compensated for their time.

At the time of writing (May 2022), over half of our participants (n=72) completed the online tasks and 32 were filmed face-to-face, completing all tasks.

3.3 Data Collection Technologies

For onsite filming, we used three high-definition digital video cameras to provide a close-up of each individual as well as one camera positioned to include both participants in the frame (see Figure 3 below). When necessary, portable studio lighting was used to ensure that the best images of the participants were captured.



Figure 3: Three angles of the cameras: one close up view of each participant, and a third camera positioned to capture both participants

In addition, two Microsoft Kinect Azure cameras were used to track the motion of participants whilst retelling the two elicited narratives. The Kinect Azure cameras consist of an RGB camera and an infrared camera (Brown Kramer et al., 2020). The system supplies a skeleton representation of the participant, consisting of X, Y, Z coordinates of 32 major skeleton joints connected by line segments. These are used to calculate a variety of movement measures such as signing speed, volume, variance, etc., which can be compared across participants, social groups, and even languages. Motion capture has also been utilized in other sign language corpus projects such as the DGS corpus project.

3.4 Data Coding

The completed data collection is estimated to consist of 360 hours of recordings (120 participants x 3hrs). The corpus will be annotated by students and research assistants, using ELAN, a video annotation software (Crasborn & Sloetjes, 2008).

4. Related Resources

The corpus will be made publicly available via the corpus **website**: www.islcorpus.co.il. The website, which is currently under construction, will follow the format of other open-access corpora, in which it will be possible to search and download the data by request. A copy of all of the elicitation materials can be found on **Open Science Framework** (<https://osf.io/yma98/>).

The corpus data will provide the dataset for future versions of an online lexical database known as **ISL-LEX**. ISL-LEX, created by the SIGN-LEX team in the US (Caselli et al., 2022) in collaboration with ISL research teams in Israel, is an online interface and search tool associated with

an existing lexical database of ISL. It provides network-based visualizations of ISL signs based on phonological characteristics, as shown in Figure 4 (<https://sites.google.com/view/isl-lex>). In its current version, ISL-LEX contains 961 signs grouped and colored by the degree of similarity to other signs (Morgan et al., 2022). In the next version, ISL-LEX will expand by using the dataset collected as part of the Corpus of ISL.

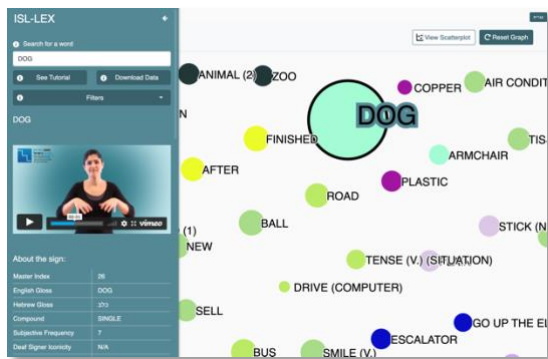


Figure 4: Example of ISL-LEX

Videos of individual signs and their ID-glosses from the corpus will be stored in an online lexical database hosted by **Global SignBank** (Crasborn et al., 2020). SignBank is a lexical database for managing ID-glosses and information about the sign form, which is dynamically-linked with ELAN for ease of coding. The database is available in three languages: English, Hebrew, and Arabic (see Figure 5 below). The goal of the multilingual format is to make the database as widely accessible as possible, and especially to different deaf communities across Israel. Data from ISL-LEX version 1.0 served as the initial input into the ISL database.

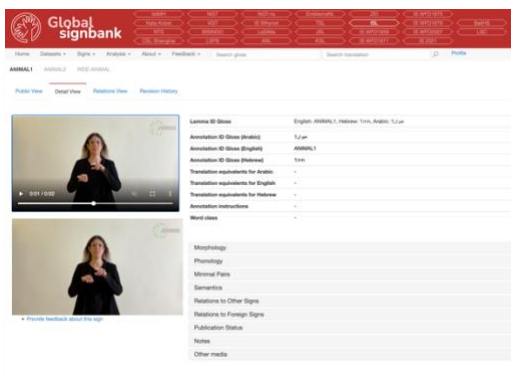


Figure 5: View of the ISL dataset as displayed in Global SignBank

5. Conclusion

In this paper, the project aims, methodologies, and the related resources were presented. The Corpus of ISL joins many other sign language corpus projects launched in the last twenty years, however, this corpus offers a unique addition by providing a corpus of a relatively young sign language. The corpus will first and foremost serve as a

resource for researchers, allowing on-going and new projects on ISL contact, variation, and change. Further to this, the corpus will provide a vital open-access resource for teachers, interpreters, students, and hearing parents of deaf children.

6. Acknowledgements

We express our gratitude to the fieldworkers, the Deaf Advisory Committee, and all the deaf participants involved in the corpus filming. We also thank Professor Adam Schembri for his support as our International Consultant. Finally, a special thanks to the SIGN-LEX and Global SignBank teams for their support and collaboration.

Funding for building the Corpus of ISL was supported by an Israeli Science Foundation grant to Dr. Rose Stamp (2757/20; ISL Corpus project) and an internal Data Sciences Institute (DSI) grant, Bar-Ilan University, awarded in 2020-2021.

7. Bibliographical References

- Börstell, C. 2017. Types and trends of name signs in the Swedish Sign Language community. *SKY Journal of Linguistics*, 30 (2017), 7-34.
- Brown Kramer J., Sabalka L., Rush B., Jones K., Nolte T. (2020). *Automated depth video monitoring for fall reduction: A case study*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; Seattle, WI, USA. pp. 294-295
- Crasborn, O., & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language. *Proceedings of LREC 2008, Sixth International Conference on Language Resources and Evaluation*. pp.39-43.
- Kopf, M., Schulder, M. & Hanke, T., (2021). Overview for datasets for the sign languages of Europe. EASIER report. University of Hamburg. <https://doi.org/10.25592/uhhfdm.9560>
- Kastner, I., Meir, I., Sandler, W., & Dachkovsky, S. (2014). The emergence of embedded structure: Insights from Kafr Qasem Sign Language. *Frontiers in Psychology*, 5, 525, pp. 1-15. <https://doi.org/10.3389/fpsyg.2014.00525>
- Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. In: P. Trudgill, D. Britain, & J. Cheshire, (Eds.). In P. Trudgill, D. Britain, & J. Cheshire (Eds.), *Social Dialectology: In Honour of Peter Trudgill*. (pp. 223-244.). John Benjamins.
- Labov, W., (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.
- Lucas, C., Bayley, R., & Valli, C. (2001). *Sociolinguistic Variation in American Sign Language*. Gallaudet University Press.
- McEnery, T., & Ostler, N. (2000). A New Agenda for Corpus Linguistics—Working with all of the World's Languages. *Literary and Linguistic Computing*, 15(4), pp. 403-419.
- McKee, R.L. & McKee, D. (1999). Name signs and identity in New Zealand sign language. In Metzger, Melanie (ed.), *Bilingualism and identity in Deaf communities*, pp. 3-40. Washington, DC: Gallaudet University Press.
- McKee, R., & McKee, D. (2011). Old Signs, New Signs, Whose Signs? Sociolinguistic Variation in the NZSL Lexicon. *Sign Language Studies*, 11(4), pp. 485-527.

- Meir, I., & Sandler, W. (2008). *A Language in Space: The Story of Israeli Sign Language*. Routledge. <https://www.routledge.com/A-Language-in-Space-The-Story-of-Israeli-Sign-Language/Meir-Sandler/p/book/9780805862652>
- Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and Interpretation*. Wiley.
- Morgan, H., Sandler, W., Stamp, R. & Novogrodsky, R. (2022). ISL-LEX v.1: An Online Lexical Resource of Israeli Sign Language. Proceedings of the LREC conference (June 25th 2022), Marseilles, France.
- Saiegh-Haddad, E., (2012). Literacy Reflexes of Arabic Diglossia. *Current Issues in Bilingualism*, no. 5: pp. 43-55.
- Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S., & Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation and Conservation* 7, pp. 136-154.
- Stamp, R., & Jaraisy, M. (2021). Language Contact between Israeli Sign Language and Kufr Qassem Sign Language. *Sign Language Studies*, 21(4), pp. 455–491. <https://doi.org/10.1353/sls.2021.0007>
- Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., & Cormier, K. (2014). Lexical Variation and Change in British Sign Language. *PLOS ONE*, 9(4), e94053. <https://doi.org/10.1371/journal.pone.0094053>
- Stamp, R., Hel-Or, H., Cohn, D., Raz, S. & Sandler, W. (2018). Kinect-ing the dots: Motion capture technologies and the sign language lexicon. Paper presented at the International Congress of Linguists (ICL) conference. University of Cape Town, South Africa. July 3rd, 2018.
- Trudgill, P. J. (1986). *Dialects in contact*. Blackwell.

8. Language Resource References

- Caselli, N., Sevcikova Sehyr, Z., Cohen-Goldberg, A. M., Emmorey, K., Pandit, S., Cao, X., Fantès, M., Simeon, J. (2022). *SIGN-LEX: Interactive visualizations for sign language lexicons*. SIGN-LEX. <https://www.sign-lex.org>
- Crasborn, O., Zwitserlood, I., van der Kooij, E., & Ormel, E. (2020). *Global Signbank*. Radboud University, Nijmegen. <https://signbank.science.ru.nl>.
- Institute for the Advancement of Deaf Persons in Israel (IADPI). *ISL Dictionary*. <https://isl.danfishgold.com/#he-3SI>
- Stamp, R., Ohanin, O. & Lanesman, S. (2022). *The Corpus of Israeli Sign Language*. Bar-Ilan University, <http://isllcorpus.co.il> (landing site launched in 2022).