IPSL: A Database of Iconicity Patterns in Sign Languages. Creation and Use

Vadim Kimmelman, Anna Klezovich, George Moroz
University of Amsterdam, Higher School of Economics, Higher School of Economics
Amsterdam (the Netherlands), Moscow (Russia), Moscow (Russia)
{vadim.kimmelman, belkannkl, agricolamz}@gmail.com

Abstract

We created the first large-scale database of signs annotated according to various parameters of iconicity. The signs represent concrete concepts in seven semantic fields in nineteen sign languages; 1542 signs in total. Each sign was annotated with respect to the type of form-image association, the presence of iconic location and movement, personification, and with respect to whether the sign depicts a salient part of the concept. We also created a website: https://sl-iconicity.shinyapps.io/iconicity_patterns/ with several visualization tools to represent the data from the database. It is possible to visualize iconic properties of separate concepts or iconic properties of semantic fields on the map of the world, and to build graphs representing iconic patterns for selected semantic fields. A preliminary analysis of the data shows that iconicity patterns vary across semantic fields and across languages. The database and the website can be used to further study a variety of theoretical questions related to iconicity in sign languages.

Keywords: iconicity, sign language, linguistic typology

1. Iconicity Patterns in Sign Languages

Iconicity is a fundamental property of human languages, both in the spoken and signed modalities (Perniss et al., 2010). Following Taub (2012), we define iconicity as a presence of a mapping (resemblance) between a mental image and the phonetic form of a sign (that is, sounds in spoken languages, or handshape, movement, and location in sign languages). For instance, the sound of the English word *ding* resemble the sound of a bell, and the handshape of the sign TREE in Russian Sign Language (RSL) (the palm with all fingers outstretched) resembles a crown of a tree—therefore, the word and the sign are iconic.

For various reasons (see Taub (2012) for a discussion), until recently, iconicity has not been actively investigated even by researchers working on sign languages. However, in recent years, a theoretical model of iconicity has been suggested (Taub, 2001), and several empirical studies have compared iconicity in different sign languages and in gestures of hearing people (Brentari et al., 2015; Padden et al., 2013; Padden et al., 2015).

Taub (2001) develops the Analogue-Building Model of Linguistic Iconicity.[1] According to this model, iconic encoding of a concept comprises of three steps: image selection (a particular image representing a concept is selected), schematization (the image is simplified), and encoding (the image is encoded using the linguistic means). For instance, to create an iconic sign for the concept 'tree', first an image of a prototypical tree is selected (a tree with a trunk and crown), then it is schematized (for instance, the schematized representation does not refer to leaves), and then it is encoded (e.g, the hand is used to represent the crown, and the arm to represent the trunk).

Based on this model, Taub (2012) discusses that iconic signs can be further classified according to the type of the concept-image association (how the image is selected to represent the concept) and the image-form association (how the phonetic shape is selected to represent the image). For instance, concerning concept-image association, sometimes only a part of the object is represented by the sign (e.g. representing the whiskers of a cat for the concept 'cat'). Concerning the image-form association, in some cases we observe *object* handshapes where the hand represents the object itself (e.g. a flat hand representing a knife, as in the Greek sign KNIFE), while in others we observe *handling* handshapes where the hand represents the hand of an agent handling an object (e.g. a fist representing a hand of someone holding a knife, as in the British sign KNIFE)).

Several recent studies focused on this latter typology, specifically, on the choice between the object and handling handshapes in iconic signs. For instance, Padden et al. (2013) compare iconic signs for instruments in American Sign Language (ASL), Al-Sayyid Bedouin Sign Language (ABSL), and New Zealand Sign Language (NZSL), as well as gestures of hearing Americans and Bedouins. They find that both groups of hearing gesturers have very strong tendency to produce handling handshapes for instruments, while ASL and ABSL signers have a strong tendency to use object handshapes in the corresponding signs; NZSL turned out to be in between in this respect. In a different study, Padden et al. (2015) find that signers of ASL but not gesturers use the object/handling difference to encode the difference between nouns and verbs. Brentari et al. (2015) compare signers and gesturers from two countries (the US and Italy) and find similarities and differences between the groups. The research so far thus indicates that (1) there is typological variation between sign languages and languages vs. gesture variation in the choice of the form-meaning mappings and (2) there are linguistic factors that influence the choice of mappings (e.g. encoding the noun-verb distinction).

However, all these studies have been based on a very small number of sign languages, they only focused on one aspect of iconicity (the choice between the object vs. handling handshapes), and only in one semantic field (namely, instruments). Another approach is represetned by the ASL-

[1] The model applies to both spoken and signed languages, but we focus on the latter in this paper.

Lex database (Caselli et al., 2016) which contains approximately 1000 ASL signs annotated (among other features) for iconicity ratings. However, the ratings only reflect a degree of iconicity (on a 7 point scale) for the whole signs, and does not discuss iconic features. Furthermore, it only contains data from one sign language. The same is true for the study in Cates et al. (2013) who annotated more than 700 ASL signs for iconicity of the three major parameters: hanshape, location, and movement. However, these parameters were only annotated as being iconic or non-iconic, without further analysis of iconicity.[2]

In order to study iconicity patterns in a more systematic manner, we created the database "Iconicity Patterns in Sign Languages" (IPSL).[3] It contains 1542 signs from seven semantic fields in nineteen sign languages and annotated them according to five iconic parameters. In addition, we created a website with several tools to visualize the iconicity patterns. In the rest of the paper we describe the creation of the database, the features of the website, and illustrate how they can be used in research on iconicity in sign languages.

2. Creating a Database of Iconicity Patterns

2.1. The Data

In order to create the IPSL database, we used the on-line dictionary of sign languages Spreadthesign (www.spreadthesign.com). This dictionary contains video recordings of isolated signs and signed sentences (up to 15 000 entries per language) in 31 sign languages. The dictionary has been created as a tool to facilitate learning of sign languages across the world; it has not been specifically designed for linguistic research. For instance, typically only one sign per concept is provided, while a sign language might have several signs. In addition, the procedure of data collection lead to the fact that lexical signs and multi-sign descriptions of non-lexicalized concepts are not systematically distinguished. However, since we focused only on the most basic concrete concepts, we consider the data to be good enough for our purposes.

We selected nineteen sign languages from all languages present at the website, namely Russian, French, American, British, Spanish, Italian, German, Polish, Brazilian, Turkish, Portuguese, Czech, Lithuanian, Swedish, Greek, Romanian, Latvian, Estonian, and Icelandic Sign Languages. The choice was governed by two main considerations: (1) we only selected the languages for which a majority of the 15 000 signs were present on the website at the moment of selection (the fall of 2016); (2) we excluded Ukranian and Belorussian Sign Language due to their close relatedness to Russian Sign Language.

We included 87 concrete concepts from seven semantic fields: transport, nature, instruments (tools), house, clothes, food, and animals. We reasoned that this selection represents a reasonable sample of the basic concrete concepts.[4]

It turned out that some sign languages are missing signs for some of the concepts. Furthermore, we decided to exclude any entries where a concept was described by a sequence of more than two signs, as they are unlikely to be lexicalized. This resulted in the database of 1542 annotated signs (out of the 1653 theoretically possible items).

2.2. Basic annotations

The annotation is based on Taub (2012) with some important modificaitons. First, we annotate each sign with respect to the form-image association (Taub, 2012). This specifically refers to the role that the handshape plays in depicting the object. We distinguish four types of form-image association:

1. Handling: the hand of the signer represents the hand of an agent that holds or handles the object, as in the RSL sign HAMMER;

2. Object: the hand depicts the shape of the object itself, as in the RSL sign CHAIR;

3. Contour: the hand represents the outline or surface of an object, as in the RSL sign HOUSE;

4. Tracing: the hands move to trace an outline or surface of an object, as in the RSL sign MOUNTAIN;

Following an insight from Taub (2012), we introduced a binary iconic feature of personification. It receives a positive value if the whole body (the hands, arms, upper body, and head) are a part of the iconic representation. A sign with a contour or tracing associations cannot involve personification by definition, because the hands do not represent the hands of a person. Signs with handling handshapes necessarily involve personification, but a sign with an object handshape can also involve personification, for instance, in the case of the hands representing paws or wings of an animal, as in the Polish Sign Language sign BIRD.

Notice that the form-meaning association feature concerns the role of the handshape. However, locations and movements in the signs can also be iconic. We thus also introduced a binary iconic feature of location. It receives a positive value if the location of the sign is iconic, that is, it represents the marked location of the object that is being represented. This happens with signs like SUN, MOON, SKY (see the RSL sign SUN) which are often located above the neutral space, or with pieces of clothing which are often located on the body (see the RSL sign T-SHIRT).

We also introduced a binary iconic feature of associated action. It receives a positive value if the movement in the sign depicts an action (independent movement or handling) associated with the object, as in the RSL sign KEY.

Finally, we decided to annotate the signs according to one of the possible concept-image associations (Taub, 2012), namely we introduced the part-whole feature, which receives a positive value if the sign represents not the whole

[2]The list of references on iconicity here is far from exhaustive, however, to our knowledge, no research comparable to the current project has ever been conducted.

[3]Note that this abbreviation is also sometimes used for Indo-Pakistani Sign Language.

[4]We did not include abstract concepts because iconic signs expressing abstract concepts contain an additional level of complexity, namely metaphorical mapping (Taub, 2001). This can be a topic of future research and extension of the database.

object, but only a noticeable part of the object, as in the RSL sign CAT.[5]

The iconicity features interact with each other in intricate ways. For instance, as discussed above, the positive value of personification is only possible for object and handling handshapes. We discuss the interactions between the features and the rules that we postulated to make the annotations consistent in detail on the About tab of our website.

The iconicity features that we selected are probably not exhaustive. However, they represent the iconic properties of all the major parameters of a sign (handshape, movement, and location). The database can be further extended with any novel features deemed necessary by future researchers.

Not all signs in the IPSL database are iconic. We distinguish two types of non-iconic signs: those involving fingerspelling, and those which do not involve fingerspelling and for which we cannot find any iconic motivation. Note that this means that the signs are in fact not transparent, but not necessarily non-iconic. We also tried to be careful in assigning iconic motivation to dubious cases, which means that in cases of doubt we would use the non-iconic label instead of trying to come up with a far-fetched interpretation.

2.3. Compounds and two-handed signs

Some signs have a complex morphological (or at least phonological) structure. We distinguish two cases that are annotated following additional rules, namely compounds and asymmetric two-handed signs.

Compounds are signs that have two clearly distinguishable sequential parts. For compounds, we annotate each part separately, using the "&" sign to separate the values of features: e.g. *handling&object* for a compound with the handling association in the first part, and the object association in the second part, as in the Italian sign SPOON.

Another case of complex signs are two-handed signs. In symmetric two-handed signs, the two hands have the same handshape and movement, so the hands are also necessarily the same with respect to the iconicity features, and we do not annotate them separately. However, in asymmetric two-handed signs the two hands might (and often do) depict separate parts or aspects of the depicted object. For instance, in the RSL sign HELICOPTER, one hand represents the cabin and the other hand the turning blades. Since these two-parts can be different with respect to iconicity features, we annotate them separately, using the "+" sign between the parts. The active hand is always annotated first: e.g. *handling+object*, see for instance the Turkish sign SPOON.

Sometimes one or both parts of a compound are asymmetric two-handed signs. This represents the most complex type of cases in our dataset, and it is annotated following the rules described above. For instance, we annotate the form-image mapping of the sign Icelandic sign LAMP as *tracing+contour&object+contour*.

[5]Taub (2012) also considers associated action to be a type of concept-image association. We are agnostic with respect to the exact nature of this feature, but still consider it a useful feature to systematically describe iconicity.

2.4. Reliability of Annotations

In order to ascertain the reliability of the annotation procedure, the whole data set (1542 sign tokens) has been fully annotated by Authors 1 and 2 independently. We then compared the annotations to calculate agreement per each of the features that we annotated.

In order to adjust the estimation for chance agreement, we calculated Cohen's kappa statistics using the *fmsb* package (Nakazawa, 2017) in R (R Development Core Team, 2016). We considered compounds and asymmetric two-handed signs as containing multiple observations, so we separated them into separate cells (e.g. *handling+object* was turned into two observations: *handling* and *object*, located in separate cells in the dataset).

The resulting estimated values of Cohen's kappa, with confidence intervals and qualitative characterizations (as provided by the *fmsb* package) for iconic features, are summarized in Table 1.

	Cohen's kappa	CI
Form-image	0.87	0.85-0.89
Personification	0.78	0.75-0.81
Associated action	0.8	0.78-0.83
Location	0.76	0.73-0.78
Part-whole	0.65	0.61-0.69

Table 1: Annotators' agreement per iconic feature

From the table it should be clear that we have high agreement for all features, especially for the form-image mapping. The lowest agreement is observed for the part-whole feature. This is explained by the fact that the authors originally interpreted this feature in a different way for three classes of signs. Firstly, the signs for animals where the hands represent paws/wings of an animal should be analyzed as [-part-whole], as the whole body represents the body of the animal. One of the authors however consistently glossed these signs as [+part-whole]. The other two classes are the asymmetrical signs discussed in the previous section, in which the role of the second hand is unclear with respect to this feature.

After the quantitative analysis of agreement, the two authors discussed the cases of disagreement and agreed upon a correct annotation for each of those cases. Therefore, the final dataset can be considered reliably annotated following the guidelines described in this paper and on the website.

3. The IPSL Website

To make the database easily accessible to other researchers interested in iconicity, we created the IPSL website: https://sl-iconicity.shinyapps.io/iconicity_patterns/, built with *shiny* (Chang et al., 2017). The website contains a detailed description of the creation of the IPSL database, the full database (which can be searched on-line with *dt* (Xie, 2016) or downloaded), and visualistion tools.

The first tool visualizes the concepts on the map of the world. The user can select one of the 84 concepts, which are repseresented as dots on the map with color-coding of the form-image mapping. Each dot on the map is clickable and opens the video of the relevant sign (see Figure 1). The user can also filter the data points by specifying the values

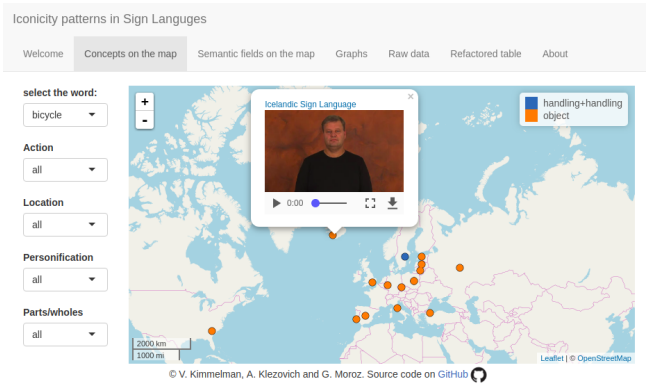

Figure 1: Concepts on the map: the sign BICYCLE.

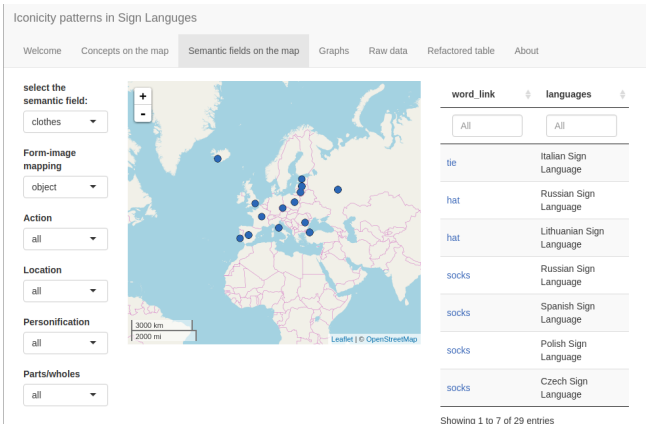

Figure 2: Semantic fields on the map: clothes.

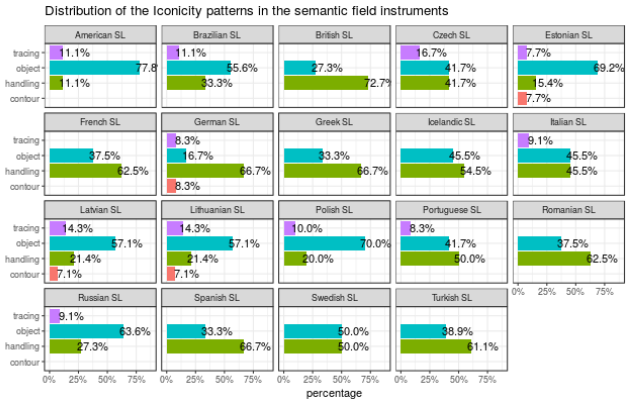

Figure 3: Form-image mappings for instruments.

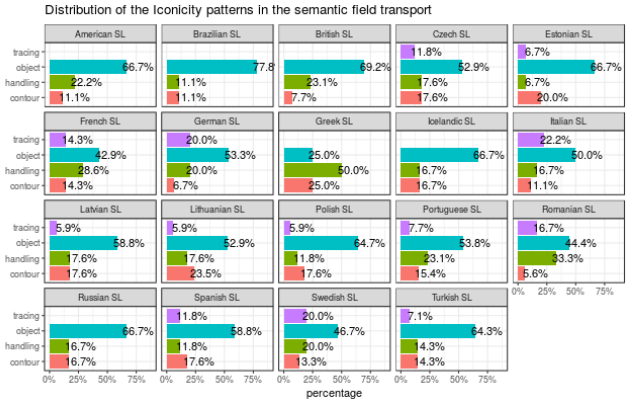

Figure 4: Form-image mappings for transport.

of the associated action, localization, personification, and part-whole features. The maps are created with the *lingtypology* package (Moroz, 2017).

The second tool visualizes the semantic fields on the map. The user can select one of the seven semantic fields, and then specify the form-image mapping and the values of the associated action, localization, personification, and part-whole features; the signs which conform the selection appear on the map (see Figure 2). In addition, a table is generated containing the concepts and languages with the relevant feature combinations.

Finally, the third tool creates graphs (bar charts) built with *ggplot2* (Wickham and Chang, 2016). The user can select a semantic field (or all semantic fields), and the charts show the distribution of concepts by the form-image mapping feature for each language, either in absolute values or in percentages (see Figure 3).

4. Linguistic Analysis

The database and the website can be used for various types of linguistic analyses of iconicity. The first conclusion that can be drawn from data is that iconicity patterns are influenced by both linguistic and semantic differences.

For instance, we can replicate the findings from Padden et al. (2013) and Brentari et al. (2015) for the form-image mappings in the domain of instruments. As Figure 3 shows, sign languages indeed vary with respect to this feature: some (e.g. ASL and RSL) prefer the object mapping, others (e.g. British and Romanian Sign Languages) prefer the handling mapping, while some (e.g. Italian Sign Language) do not show a strong preference.

Another finding not previously studied in detail is that different semantic fields have different preferences. For instance, while in the field of instruments the object and handling mappings are prevalent, for transport all but one sign language use predominantly the object mapping (Figure 4). Other iconic features also show some cross-linguistic variation and semantic preferences. For instance, Romanian Sign Language seems to use iconic localization relatively less frequently than other sign languages, while ASL has the highest proportion of signs that uses parts of the object to represent the object. Semantically, iconic locations are used predominantly in nature-related and clothing-related concepts, while personification, in addition to being used in all signs with the handling mapping, is commonly used in signs for animals.

5. Future Use of the IPSL Database

There is a variety of ways in which the IPSL database can be used for further research.

We are currently investigating a number of questions based on the database (Kimmelman et al., in preparation). Firstly, we are ivestigating whether the language and the semantic fields are indeed significant predictors of all iconicity features, and whether these predictors interact with each other. Secondly, we are studying the interactions of the iconicity features in order to find patterns and frequent constellations of features which might reflect some common iconic strate-

gies. Finally, we are comparing the sign languages to each other with repsect to iconicity to see whether such a measurement would reflect historical and geographical connections between them.

The database might be further extended. It is clear that the list of iconic features is not exhaustive: for instance, we have not considered non-manual expressions. The analysis of iconicity in terms of handshape, location, and movement might be too rough: maybe these parameters should be further subdivided and analysed for iconicity in more detail.

In addition, the database can be extended by adding more signs: extra signs for the existing semantic fields, concrete signs from other semantic fields, and abstract signs. It is clear that 87 concepts are not enough to generalize our findings to the whole lexicon.

Another possible extension would be an additon of signs for abstract concepts. As demonstrated by (Taub, 2001), abstract concepts are expressed in sign languages primarily through metaphoric-iconic signs. This means that an abstract concept is associated via metaphoric mapping with a concrete image, and this image is then enocded iconically. For instance, a sign for FEAR might be based on a metaphorical mapping between fear and the heart beating fast, and the concrete event of the heart beating fast is then encoded iconically (as e.g. in Austrian Sign Language: FEAR). If such signs are included in the database, we can annotate the iconic features for the concrete concept the usual way, but novel features describing the nature of metaphoric mapping should be added.

It would also be interesting to investigate the relationship between iconic features identified here and lexical features of these signs. This can be done by annotating further lexical information for the signs in the databse. For instance, these lexemes can be analyzed in terms of Pustejovsky's qualia structure (Pustejovsky, 1995), to find out whether certain iconic features of lexemes correspond to certain sub-features of their qualia structures.

To sum up, in this paper, we reported the process of creation of the IPSL database of iconic features in 1542 signs from nineteen sign languages and of the database website containing several visualization tools. We also demonstrated how the database can be use to replicate and extend previous typological approaches to iconicity in sign languages, and formulated a number of research questions that can be further investigated with the database.

6. Acknowledgements

The publication was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2018 (grant Nr. 18-05-0053) and by the Russian Academic Excellence Project "5-100". Vadim Kimmelman's research has been supported by the the Dutch Science Foundation (NWO) grant Nr. 360-70-520.

7. Bibliographical References

Brentari, D., Renzo, A. D., Keane, J., and Volterra, V. (2015). Cognitive, Cultural, and Linguistic Sources of a Handshape Distinction Expressing Agentivity. *Topics in Cognitive Science*, 7(1):95–123, January.

Caselli, N. K., Sevcikova Sehyr, Z., Cohen-Goldberg, A. M., and Emmorey, K. (2016). ASL-Lex: A lexical database for ASL. *Behavior Research Methods*, 49(2):784–801.

Cates, D., Gutiérrez, E., Hafer, S., Barett, R., and Corina, D. (2013). Location, location, location. *Sign Language Studies*, 7(1):433–461.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J., (2017). *shiny: Web Application Framework for R*.

Kimmelman, V., Klezovich, A., and Moroz, G. (in preparation). Iconicity Patterns in Sign Languages. Manuscript, Moscow-Amsterdam.

Moroz, G., (2017). *lingtypology: easy mapping for Linguistic Typology*.

Nakazawa, M., (2017). *fmsb: Functions for Medical Statistics Book with some Demographic Data*.

Padden, C., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., and Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture*, 13(3):287–308.

Padden, C., Hwang, S.-O., Lepic, R., and Seegers, S. (2015). Tools for Language: Patterned Iconicity in Sign Language Nouns and Verbs. *Topics in Cognitive Science*, 7(1):81–94.

Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, 1.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT press, Cambridge.

R Development Core Team, (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge University Press, Cambridge.

Taub, S. F. (2012). Iconicity and metaphor. In Roland Pfau, et al., editors, *Sign Language: An International Handbook*, pages 388–412. De Gruyter Mouton, Berlin.

Wickham, H. and Chang, W., (2016). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.

Xie, Y., (2016). *DT: A Wrapper of the JavaScript Library 'DataTables'*.