Towards an RDF Representation of the Infrastructure consisting in using WordNet(s) as a conceptual Interlingua between multilingual SignLanguage Datasets

Thierry Declerck<sup>1</sup>, Thomas Troelsgård<sup>2</sup> and Sussi Olsen<sup>3</sup> <sup>1</sup>German Research Center for Artificial Intelligence (DFKI GmbH) <sup>2</sup>University College Copenhagen (KP) <sup>3</sup>Centre for Language Technology,NorS, University of Copenhagen Goal: Representing and Publishing Sign Language Data Sets in the Linguistic Linked Data (LLOD) Cloud, which is a subset of the Linked Data (LD) Cloud





## Current status of the relation between Linked Data and Sign Languages

- We could observe that Sign Language (SL) lexical data are not represented in the datasets included by now in the LLOD cloud.
- Also looking at the "Overview of Datasets for the Sign Languages of Europe" published by the "Easier" European project (https://www.project-easier.eu/wpcontent/uploads/sites/67/2021/08/EASIER-D6.1-Overview-of-Datasets-for-the-Sign-Languages-of-Europe.pdf) we do not see any mention of a dataset being available in an Linked Data compliant format.

Checking the usability of Ontolex-Lemon for encoding Sign Language data – Ontolex-Lemon covers only writtenRep and phoneticRep. How to include the Representation of SLs?



The Features of Sign Language Data we want to integrate in the LLOD cloud? An example from the American Sign Language



Figure 2: Various representations of American Sign Language. English translation: "What is your name?"

Taken from: Including Signed Languages in Natural Language Processing. Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, Malihe Alikhani. https://arxiv.org/abs/2105.05222 (<u>ACL-IJCNLP 2021</u>)

## About the Phonology of Sign Languages

"Sign languages such as <u>American Sign Language</u> (ASL) are characterized by <u>phonological</u> processes analogous to, yet dissimilar from, those of oral <u>languages</u>. Although there is a qualitative difference from oral languages in that sign-language <u>phonemes</u> are not based on sound, and are spatial in addition to being temporal, they fulfill the same role as phonemes in oral languages.

Basically, three types of signs are distinguished: one-handed signs, symmetric two-handed signs (i.e. signs in which both hands are active and perform the same or a similar action), and asymmetric two-handed signs (i.e. signs in which one hand is active [the 'dominant' or 'strong' hand] and one hand is held static [the 'non-dominant' or 'weak' hand]). The non-dominant hand in asymmetric signs often functions as the location of the sign. Almost all simple signs in ASL are monosyllabic."

(https://en.wikipedia.org/wiki/American Sign Language phonology)

More details in specialized literature.

## A first Step for encoding the Phonology: A (first and tentative) Ontology of constitutive Elements of SLs

💠 TBCFreeWorkspace - TopBraid - TopBraid Composer FE				– 🗆 X
File Edit Navigate Model System Inference Resource Window Help				
📑 🕶 📰 🕼 😫 🌭 💽 🕶 🛤 🔝 🛣 🔛 🏠 🐃 🗢 🕶 😑 sldch	lovement	A     A		Quick Access 🖹 😭 🔂 🗞
🍃 Classes 🛛 👘 🔽 🐨 🗖 🗖	lexinfo.ttl 🛛 🗟 SignLar	igDatCat.ttl 🛛 🗖 🗖	🎦 Properties 🔀	📫 😨 🗸 🗖 🗖
<pre>keyse keyse k</pre>	Class Form URL: http://example.org/sig Vanotations dcsource v Si https://datcatinfo.termw rdfstabel v skossdefinition v skossdefinition v Class Axioms rdfssabclassOf v	G     G     III     ▼       n_language_data_cat#Mov       eb.eu/search/terms     ▼       a sign." (@en)     ▼	Chttp://www.w3.or     Chttp://wwww.w3.or     Chttp://www.w3.or     Chttp://www.w3.o	g/ns/lemon/lime≢onceptualD≥ ∧ g/ns/lemon/lime≢entry> g/ns/lemon/lime≢lexiconDatas g/ns/lemon/lime≢referenceData g/ns/lemon/lime g/ns/l
> 🥚 sldc:NotationsForSignLanguages (9)	sldc:Phonology_Manual		ontolex:concept	
<ul> <li>✓ ● sldc:Sign (254)</li> <li>&gt; ● sldc:Duration</li> <li>&gt; ● sldc:Fingerspelling</li> </ul>	owl:equivalentClass ▽ owl:disjointWith ▽		ontolex:evokes ontolex:isConcept ontolex:isDenotedE	Df By
> 🛑 sldc:LexicalProperties (15)	owl:hasKey ▽		ontolex:isEvokedBy	Y
✓ sldc:Phonology (239)	Other Properties	Y	<	>
sldc:Phonology_Manual (210)	Form Source Code			* 8
side:Dominantriand	🧐 Imports 🔶 Instances 🛛	🔳 Domain 🥺 Error Log 关	SPARQL 🔗 Text Search	🗳 🍖 🗢 🗖
<ul> <li>sldc:NandetivatedFingers (1)</li> </ul>	[Resource]	rdf:type	rdfs:label	rdfs:comment
sldc:SelectedFingers (12) sldc:ThumbContact (2)	sldc:ArticulatorInternal	sldc:ArticulatorInternalMo	"kinking changes in hands	DGS
sldc:ThumbPosition (2)	<ul> <li>sldc:ArticulatorInternal</li> </ul>	sldc:ArticulatorInternalMo	"Opening/Closing change	DGS
🗸 🛑 sldc:Location (105)	<ul> <li>sldc:Articulatorinternal</li> </ul>	side: Articulator Internal Mo	"Rubbing changes in hand	DGS
sldc:Location_DGS_Horizontal (6)	<ul> <li>sldc:ArticulatorInternal</li> </ul>	side:ArticulatorInternalMo	"bending movement"	003
sidc:Location_DGS_Vertical (22)	<ul> <li>sldc:ArticulatorInternal</li> </ul>	sldc:ArticulatorInternalMo	"rotating movement"	
side:MajorEccation (7)	sldc:ArticulatorInternal	sldc:ArticulatorInternalMo	"shaking movement"	
side:SecondMinorLocation (31)	sldc:BackAndForth	sldc:PathMovement	"back and forth"	
> 🗘 sldc:Movement (23)	sldc:Bend	sldc:PathMovement_Shape	"bend"	
sldc:NonDominantHand	sldc:Circular	sldc:PathMovement_Shape	"circular"	
> 😑 sldc:Orientation (6)	sldc:Curved	sldc:PathMovement_Shape	"curved"	
> sldc:TwoHandedSigns (3)	sldc:False	sldc:FlexionChange, sldc:l	"false"	
sldc:Phonology_NonManual (7)	sldc:None	sldc:PathMovement	"none"	
sidc:Phonology_loBeClassified (22)	sldc:Other	sldc:MajorLocation, sldc:	"other"	
	<ul> <li>sldc:OtherAway</li> </ul>	sldc:MajorLocation, sldc:	"other away"	
• location	<ul> <li>sidc:Spiral</li> <li>sidc:Spiral</li> </ul>	sldc:PathMovement_Shape	"spiral"	
•• **	<ul> <li>slucistraight</li> <li>sldciTrue</li> </ul>	sluceratniviovement_shape	suaight "true"	
🎦 Project Explorer 🛛 🚼 Basket 🛛 🖃 🔄 🤝 🗖 🔂	<ul> <li>sldc:Waves</li> </ul>	side:PathMovement Shape	"wavec"	
> 🗁 36_Dimensions	sldc:X-Shaped	side:PathMovement_Shape	"x-shaped"	
> 🗁 Eurolan21	sldc:Z-Shaped	sldc:PathMovement Shape	"z-shaped"	
> Composition Morphologies			<b>F</b> = = = =	~

See also Thierry Declerck Towards a new Ontology for Sign Languages. LREC 2022

# About the status of "Glosses" used for labelling Sign Language Data

- "Glosses" used in Sign Language resources are not originated in lexical entries of the spoken language or from specific vocabularies. More a kind of informal annotation of a performed sign:
  - "In ASL it is an English word or words that we use to name ASL signs so that we can talk about these signs. The word or words associated with that sign do not relay the signs meaning." (Signs & their Glosses - Bellevue College)

## The "GLOSSING" pages of IDGS point to Corpora where the signs are occurring, but also to more detailed phonological and lexical information

#### Pointing to corpus attestations

ABWEHR1

Frankfurt | dgskorpus\_fra\_06 | 31-45f Sie wollten mich mit Gewalt ausziehen, aber ich habe mich dagegen gewehrt.

	AUF-	AUSZIEHEN- UNTERKÖRPER1A*		ICH1	ABWEHR1
	FERSONT				
	auf	ausziehen		ich	
Mün	ister   dgskorpus	_mst_16 61+m	Dał	nabe ich mie	ch gewehrt.
				ICH1*	ABWEHR1
				5 · · - 2	6 <b>.</b> .
m				[MG]	[MG]
Mün	ister   dgskorpus	_mst_16 61+m	Ich	habe mich g	gewehrt und gewe
				ICH1	ABWEHR1
				[MG]	wehren wehren

Münster | dgskorpus\_mst\_16 | 61+m Ich habe mich gewehrt und an meiner Mutter festgehalten und gerufer "Mama, Mama."

BWEHR1 FESTHALTEN2**	ABWEHR1
[MG] mama mama	[MG]

Pointing to more detailed information (transcription, phonology, and lexical)



### Pointing to more detailed information (transcription/notation, phonology, and lexical) https://www.sign-lang.uni-hamburg.de/galex/glossen/g13990.html

#### ← SCHUTZ1A →



Pointing also to signs having a similar had configuration



### About the Status of Transcriptions

- We focus on HamNoSys ("Hamburg Sign Language Notation System"), which is a phonetic transcription system.

![](_page_10_Figure_3.jpeg)

HamNoSys or ist SiGML transcription can be encoded as a value of ontolex:Form

# Linking values of ontolex:Form to instances of ontolex:LexicalConcept

- Current (very promising) work is dealing with linking SL data to WordNet data included in the Open Multilingual Wordnet infrastructure.
  - Contribution on NexusLinguarum (and the projects SignOn and Easier), in cooperation with the University of Hamburg, the Institute for Language and Speech Processing (Greece), and the Fran Ramovš Institute for the Slovene Language: offering a Linked Data compliant representation of this type of linking: ontolex:Form to ontolex:LexicalConcept (where WordNet data is encoded), especially for the many cases for which we are lacking a lexical entry or a lexical sense.

# The Linking of SL data and OMW (for English and Greek), as proposed in the EASIER project (https://www.fdr.uni-hamburg.de/record/10169#.Y01WXExBzmE)

add new synset						
omw.01115162-n omw link internal link	dgs9292 <u>ilex link</u>					
lemmas:sell	switch to german					
def:the activity of persuading someone to buy validated: 1 confidence: 5 GSL 🗸	TO-DISTRIBU	JTE2^				
validated as correct undo validation		"⊂	<sup>2</sup> <u>r</u> 0 <sup>X[∠</sup> → <u>1</u> ]			
omw.02242464-v omw link internal link lemmas:sell		Ma	MA			
def:exchange or deliver for money or its equivalent						
validated: 1 confidence: 5	frontal 45° 90° from abov	re				
validated as correct during undo validation	TO-SELL1					
omw.00767635-v	Berlin   dgskorpus_ber_02   4	16-60m There is a c	lothes store in the c	apital of Tunisia,	no, in the second larg	jest city
omw link internal link		DFAF1A*	STORF1*	TO-SELL1	CLOTHES1A*	
lemmas:sell	\$INDEX1	DEG IN	OTORET	TO OLLET	CLOTHLO III	
def:persuade somebody to accept something	m		geschäft			
validated: 0						

The Linking between SL Data and OMW in an Excel File (for Greek and English) – Cross-Lingual Linking via WordNet IDs Conceptually, more precise then linking via Glosses

#### The Greek data

970 gsl326,omw.00988028-v,manual accept
971 gsl1049,omw.00362103-n,manual accept
972 gsl1050,omw.00377364-n,manual accept
973 gsl1050,omw.07308563-n,manual accept
974 gsl2592,omw.05128519-n,manual accept

#### The English data

 6833
 dgs67339,omw.00568430-n,auto accept

 6834
 dgs10875,omw.14449405-n,auto accept

 6835
 dgs10040,om w.00377364-n,manual accept

 6836
 dgs10481,omw.00377364-n,manual accept

 6837
 dgs9882,omw.04228054-n,auto accept

 6838
 dgs73480,omw.07349299-n,auto accept

(from https://www.fdr.uni-hamburg.de/record/10169#.Y01WXExBzmE)

#### The entry 00377364-n in the Greek Wordnet

00377364-nell:lemma έκρηξη00377364-nell:lemma σκάσιμο00377364-nell:def0φ κρήγνυται ή σκάει κάτιο

#### The entry 00377364-n in the English Wordnet

00377364-n	lemma	explosion
00377364-n	lemma	burst

Another IDGS Resource: Dicta-Sign (<u>https://www.sign-lang.uni-hamburg.de/dicta-sign/portal/concepts/cs/cs\_688.html</u>): 1000 concepts with Sign Realizations in 4 languages, using single Glosses and other videos

![](_page_14_Picture_1.jpeg)

### "Beskyttelse" in the Danish SL Portal

![](_page_15_Figure_1.jpeg)

### Transcriptions of Danish SL

#### FORSVARE • VAR

#### 

<sigml><hns\_sign gloss='FORSVARE'><hamnosys\_manual><hamsymmlr/><hamfist/><ha mparbegin/><hamextfingeru/><hampalmd/><hamplus/><hamextfinge rr/><hampalmr/><hamparend/><hamparbegin/><hammoveu/><hamt humbside/><hamtouch/><hamplus/><hamnomotion/><hamparend/>< hamrepeatfromstart/></hamnosys\_manual></hns\_sign></sigml> https://www.tegnsprog.dk/video/t/t\_2162.mp4

## DanNet – PWN core (taken from <u>https://github.com/omwn/omw-data/blob/main/wns/dan/wn-data-dan.tab</u>

4534	13244109-n	lemma	besiddelse	
4535	13244109-n	lemma	eje	ODT
4536	13244109-n	lemma	ejendom	http
4537	05091194-n	lemma	proportion	resu
4538	07161429-n	lemma	forslag	
4539	14489699-n	lemma	velstand	
4540	01128193-v	lemma	beskytte	
4541	01128193-v	lemma	forsvare	
4542	01128193-v	lemma	værne	
4543	00817680-n 💻	lemma	beskyttelse	
4544	00817680-n	lemma	forsore	
4545	00817680-n	lemma	forsvar	
4546	00817680-n	lemma	værn	
4547	14539960-n	lemma	ly	
4548	14539960-n	lemma	skjul	

DT-ID 3844:

https://www.tegnsprog.dk/#%7Csoeg%7C'tekst'beskyttelse%7C resultat%7C1%7Ctrestjerner%7C1%7Ctegn%7C837

> Signs/Videos/Glossesof Signs,Transcriptions of Signs In other Sign Languages, like English, French, German, Greek, etc..

## Our RDF/OntoLex-Lemon Encoding of the conceptual Information

Classes 🔀	- 😵 🌄 🚫 🗢 🗆	💿 sl-23-10.ttl 🛛		🚦 Properties 🛛 📫 🙀 🗍
🛑 frac:Corpus (1)	-	<ul> <li>Annotations</li> </ul>	^	🔲 dgs:hasGLOSS
🛑 frac:Observable		rdfs:label ▽		🔲 dgs:hasOMW-Link
> 🛑 geo:SpatialThing		Schutz" {@de}	$\sim$	🔲 dgs:hasSubType
lexicog:FormRestriction		== "beskyttelse" {@da}	$\bigtriangledown$	🔲 dgs:hasVideo
Iexicog:LexicographicCor	mponent	protection" {@fr}	$\overline{\nabla}$	geo:location
lexicog:UsageExample		Protection" (@en)	$\overline{\nabla}$	Iexicog:describes
ontolex:Form (5)			$\overline{}$	Iexicog:restrictedTo
> _ ontolex:LexicalEntry (9)				Iexicog:usageExample
owl:NamedIndividual		skos:definition 🗢		lime:conceptualDataset
> semiotics:Expression (9)		🔚 "παρεχόμενη φροντίδα σε κάποιον ώστε να προφυλάσσεται από υπαρκτούς ή διάφορους πιθανούς	$\overline{}$	lime:entry
> semiotics:Meaning (13)		κινδύνους" {@el}		lime:lexiconDataset
> skos:Collection		<ul> <li>Other Properties</li> </ul>		
> skos:Concept (7)				lime:referenceDataset
> skos:ConceptScheme (3	)		<b>—</b>	lime:resourceType
sl:GLOSS (8)		Deskyttelse {@da}	_	ontolex:concept
sl:SignVideos (6)		== "forsorg" {@da}	~	ontolex:evokes
synsem:OntoMap		== "forsvar" {@da}	$\overline{}$	ontolex:isConceptOf
synsem:SyntacticArgume	ent	"protection" {@en}	$\overline{}$	
synsem:SyntacticFrame		"protection" {@fr}	$\overline{}$	
	>	∎■ "værn" {@da}	$\bigtriangledown$	ontolex:IsLexicalizedSenseOf
	~ <i>/</i>	🔚 "προστασία" {@el}	$\overline{\nabla}$	
Project Evalerer 😚 👥 Packa	+	Form Source Code	•	
	- E 2 H -	📲 Imports 🔶 Instances 💻 Domain 🥙 Error Log 🔀 🧏 SPAROL 🔗 Text Search		

## Our RDF/OntoLex-Lemon Encoding of the SL Glosses

![](_page_19_Picture_1.jpeg)

## Our RDF/OntoLex-Lemon Encoding of the Locations of the SL Videos

ses 🗤 🤟 🐄 📉		
frac:Corpus (1)		dgs:hasGLOSS
frac:Observable	Resource Form	🔁 🔜 💷 🎽 🔲 🗖 🗖 🗖 🗖
geo:SpatialThing	Name: dts:SignVideos_dts-t_2162.mp4	🔲 dgs:hasSubType
lexicog:FormRestriction		🔲 dgs:hasVideo
lexicog:LexicographicComponent	Annotations	🔲 geo:location
lexicog:UsageExample	rdfs:label ♡	Iexicog:describes
ontolex:Form (5)	S "Videos representing a sign"	✓ I lexicog:restrictedTo
ontolex:LexicalEntry (9)	- Other Bronauties	🔲 lexicog:usageExample
owl:NamedIndividual	• Other Properties	🔲 lime:conceptualDataset
semiotics:Expression (9)	dgs:hasGLOSS ▽	lime:entry
semiotics:Meaning (13)	dts:GLOSS_dts-1_2162	🗢 🔲 lime:lexiconDataset
skos:Collection	sl:hasVideoAdresss ▽	🔲 lime:linguisticCatalog
skos:Concept (7)	H https://www.teansprog.dk/video/t/t_2162.mp4	
skos:ConceptScheme (3)		🔲 lime:resourceType
sl:GLOSS (8)	rdf:type ▽	🔲 ontolex:concept
sl:SignVideos (6)	sl:SignVideos	🗢 🗖 ontolex:evokes
synsem:OntoMap	owl:topDataProperty ▽	ontolex:isConceptOf
synsem:SyntacticArgument		🔲 ontolex:isDenotedBy
synsem:SyntacticFrame	<ul> <li>Incoming References</li> </ul>	🔲 ontolex:isEvokedBy
······································	Č dgs:hasVideo ▽	ontolex:isLexicalizedSenseOf
	◆ dts:Form_1_2162	□ ontolex:isReferenceOf

### Conclusions

- Promising on-going work with an exciting cross-disciplinary cooperation!
- The linking of SLs data to Wordnet seems to be an optimal anchor point for linking SL Data to spoken language data.
- Integrating a number of disparate SL, conceptual and lexical resources under one "umbrella", realising a dense linking of this distinct information sources.
- Current work in applying the approach across Nordic Sign Languages (Danish, Swedish, Icelandic for the time being)

### Thanks for your Attention!