Organizing data in a multilingual observatory with written and signed languages.

Cédric Moreau(**), Bruno Mascret(*)

Université de Lyon *, CNRS
Université Lyon 1, LIRIS, UMR5205, F-69622, France

INS HEA **
Institut National Supérieur de Formation et de Recherche pour
l'Education des Jeunes Handicapés et les Enseignements Adaptés.
58, 60 av des Landes 92150 Suresnes France

UMR 7023 **
Université Paris 8
2 rue de la Libert
93200 Saint-Denis France

cedric.moreau@inshea.fr, bruno.mascret@liris.cnrs.fr

Abstract

The Acadmie Française institution is assigned and devoted to defending the French language and to making it a common heritage for all French speakers. The French Sign Language (LSF) has never had such a support.

To face this situation, a reference tool has been created, supported by the French Ministry of Education and by the General Delegation to the French language and to languages of France. This tool is a collaborative website entirely bilingual French and LSF, and which proposes for each concept at least one definition and its associated descriptors in various knowledge fields. Before being spread on-line, the information given by users (text, picture, video, presentation) is examined by experts on form and content, and is validated or rejected by these experts with an explanation.

Considering regional and sociological differences, several signs may be proposed and validated for one concept. Our project does not wish to choose the "ideal sign", but wants to submit to our identified users all the proposals and to list their comments (have they come across this sign and if so, in which context). A set of information is thus collected for each sign and can be related to users profiles. The website is therefore an exchange platform, but can also be used as a linguistics observatory.

One of our main issues concerning the data organization was to manage to adjust users different viewpoints and different uses of the website. Indeed, our platforms goal is not to make a simple dictionary but to create a network of ontologies. Our other issue is now that we cannot use a rigid organization model, because our website must constantly evolve and include new concepts and new descriptions or functionalities, such as illustrations, homonyms, antonyms, etc. In this article we will first briefly describe our platforms goals, then present our specific data organization which allows for example several classifications to be used simultaneously. We will illustrate this approachs interest with a critic of Deweys classification, that we had at first implemented despite its limits (acceding to a precise concept is difficult, the organization is not intuitive, recent concepts or specific LSF concepts cannot be referenced, etc.). We will propose to replace it with classifications directly created by our users and corresponding to their expectations and needs. This way the tree diagram is built gradually and supervised by experts in each knowledge field.

Each content thus goes with descriptors and classifiers allowing it to play different parts depending on the context. Therefore a content can at the same time be a concept, a classification theme or sub-theme, or an illustration – the context will mobilize the appropriate contents depending on their descriptors and classifiers.

We will finally present our current work on integrating direct resources in LSF through descriptors defining a sign's spatial position and its moves (hands, body and face), in order to highlight our platforms great ability to evolve. We will also show that this data organization allows an easy conversion to other countries sign languages.

Key words: French Sign Language, LSF, written languages, dynamic classification, deaf, collaborative website, concept, ontology.

1. Preamble

According to Gillots official report[1], 80% of the French deaf people are illiterate, and only 5% reach higher education. Dalle has also declared in (Dalle, 2003) that "illiteracy, short knowledge of written French, lack of diploma and of qualifications as well as communicating problems have great consequences on deaf adults' social and professional integration".

Since the French 11th February 2005 act, public schools cannot refuse for any reason to take in a child living in its defined area. Besides, all public websites must be entirely accessible. In 2010, a state diploma will be created for LSF teachers. In this context, new needs have appeared, and appropriate bilingual teaching tools are increasingly demanded.

Just as the French language, LSF has many regional vocabulary differences, and it constantly enriches itself with new words, thanks to its speakers who create signs to name new

---

[1] http://cis.gouv.fr/spip.php?article1516

concepts and/or concepts that are specific to a knowledge field. As the deaf people increasingly access university and professional environments, this phenomenon is enhanced. (Duquesne-Belfais, 2007) stresses out that each concept – once its has been attributed a name or a sign – can be used to define a more abstract concepts characteristics and can take part in building a knowledge network. Nominalizing the concepts characteristics allows it to change its status, switching from implicit to explicit, and to take part in constructing a rigorous language.

2. "OCELLES" PROJECT

The main support of the "OCELLES" project (Moreau, 2008) is a collaborative website, entirely bilingual French/LSF, and which proposes for each concept at least one definition and its associated descriptors – in both languages – in all possible knowledge fields. Before being spread on-line, the information given by users (text, picture, video, presentation) is examined by experts on form and content, and is validated or rejected by these experts with an explanation.

The project is currently under testing and will be published at http://www.ocelles.fr.

2.1. Managing users and rights

Running with a GPL licence[2], the website is free to all. Users may access different statuses:

- visitors browse on the websites public content.

- they become users when registering and filling a form establishing their profile. The collected information – on their scholarship, track record, languages used within family and social lives – will balance their answers and advice concerning proposed signs.

- writers are users who have accepted the publication terms. They may propose new contents and concepts to experts – possibly supported by a classification. They may also add videos to illustrate other authors text sequences.

- experts are writers who validate the contents deposited by writers. They must also provide an explanation to the writer in the event of non validation.

- the administrator is expert in all fields. However his main role is managing the portal without taking part in expertises and publications.

We must specify that writer and expert statuses are only attributed to a users knowledge field: a user may be expert in mathematics, writer in philosophy and plain user in all other fields.

Managing users rights so precisely should lead to a democratic and community use of our website, because the hierarchy only depends on the chosen theme.

---

[2]http://www.gnu.org/licenses/licenses.html

2.2. Concepts

For each concept, a specific and dynamic webpage proposes a definition, a translation into written language and one or several signs in Sign Language. Considering regional and sociological differences, several signs may be proposed for each concept. Our project does not wish to choose the "ideal sign", but wants to submit all proposals to users. The definitions respect the following rules: Written definitions must:

- give the concepts meaning and its main characteristics,

- be precise,

- begin with a general explanation,

- be a suitable substitute to the unknown word in a text,

- not integrate other definitions,

- not contains other words having the same root,

- not be circular.

Illustrations, examples, comments (educational or linguistic etc.), slight differences, regional uses, connotations and other are in addiction to the definition, not a part of it. A Sign Language definition:

- must not contain any regional code or name (however local signs may be used to refer to a concept),

- must avoid neologisms,

- should be punctuated in order to enable its sequencing.

The signing flow is adjusted for deaf or hearing learners.

Figure 1: Screenshot of a "concept page".

The platforms goal is not to provide a plain dictionary, but a real network of ontologies. Links enable associations between concepts – ex. "thesis" refers to "arguing" – thanks to the "see also", "close concepts", and "opposite concepts" descriptors. The links between concepts are flexible, and more type of links can easily be integrated to "OCELLES".

2.3. signs

Each sign proposal opens a new web page. Examples, context and other descriptors as well as linguistic and epistemological comments can be added. Users are encouraged to answer questions about:

- the context(s) in which they have encountered each sign (class, job etc.),

- the sign characteristics – i.e. formed with one or two hands, coming from a transfer (Cuxac, 2000) in form, situation, person, configuration, position, moves, facial expression etc.

2.4. Proposals summary

For each concept, the answers given by other users are gathered and summed up on one page. They can be linked to their profiles, thus enabling for example a collection of information about geographical localization of each sign.

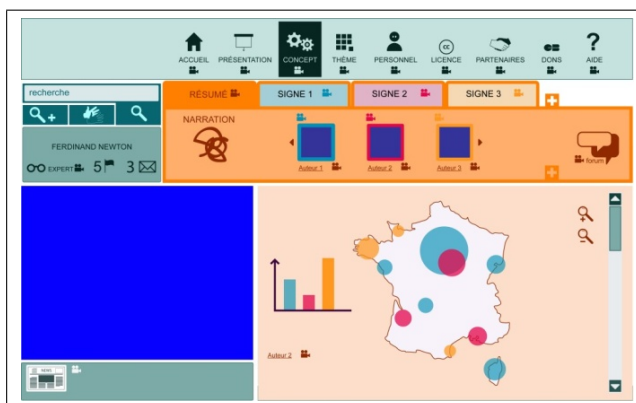More than an exchange platform, the website is also a synchronic linguistics observatory.

Figure 2: Screenshot of a "summary page".

3. Data organization model

3.1. A rigid classification and its limitations

At first we have used Deweys decimal classification (DDC), usually used in libraries. We have chosen this system, developed by Melvil Dewey in 1876[3], because it exists and classifies the whole set of human knowledge.

However, user tests, made on both deaf and hearing persons, rapidly pointed up the difficulties we had sensed. The DDC consists in classifying works and knowledge into 10 general categories, each one of them being divided and subdivided each time into 10 subcategories and so on as many times as needed. Looking for a specific concept through this tree diagram makes it imperative to:

- know into which categories and subcategories the concept will classified,

- make no mistake through the tree diagram in choosing the subdivisions.

_____
[3] http://www.oclc.org/dewey/

This approach, not intuitive, will locate a low-level concept very far from the head of the diagram, thus the low-level concepts will be found only if the high-level ones are known and understood.

As for the linguistic system LSF, it combines a categorical aspect with its vocabulary structure. According to (Courtin, 1998), the use of LSF by deaf children whose parents are deaf and already signers increases categorizing abilities compared to hearing or oralizing children. Courtin has observed this phenomenon especially when the categorization respects our worlds complexity by using prototypes or diagrams (Bideaud et al., 1993). Indeed, signers often refer to a concept through a series of prototypical examples of it, thus defining the concept by extension. A rigid and arbitrary classification could then disturb deaf users.

Besides, where and how should new concepts directly stemming from Sign Language be classified, in a rigid classification set upon written language ? (ex. "LS Video", video recordings of formalised LSF used as a differed communication, or "signary", set of all signs in Sign Language).

3.2. Dynamic classification

One of our main issues concerning the data organization is to manage to adjust users different viewpoints and different uses of the website. Our portal must be able to easily evolve and include new concepts as well as new descriptions or functionalities, such as illustrations, homonyms, antonyms, etc. Keeping using a rigid data organization is impossible.

That is why we have chosen a data organization which considers a priori each one of the web site's elements as a content. In parallel, an associative and dynamic structure has been set up, enabling to link contents together according to their roles and to the descriptors associated to these roles (Moreau and Mascret, 2008).

This way, one content may be used several times because in different contexts, depending on the associations it belongs to – roles and descriptors (Bénel, 2003).

Let us give an example : in our website, a classification node has a role of theme. A theme is also a concept for example "language". This theme lists other themes and concepts. "Language", as a theme, contains the themes "lexicon", "grammar"... Moreover, "language", as a content, also has a role of illustrator to the concept "lexicon". This way, one content – here "language" – has different roles (theme, concept). Each one of these roles has its own descriptors (concept's illustration, other related concept ...).

Finally, the diagram tree must allow a concept to be classified in several themes without duplicating it. Libraries often face this problem when classifying works containing several themes – a book about science in the 19th century should be classified into history as well as science.

3.3. Discussion : a dynamic classification built on LSF linguistic parameters?

According to (Cuxac, 2000), two discursive enunciation strategies coexist in LSF. Through the visual-gestual channel, the signer chooses to say without showing, or to say and show. This way, he can visually re-present the experience thanks to the greatest resemblance between a sequence of signs and the experience itself. Or else he can use the

standard sign that does not resemble the referent. Based on this theory of iconicity, our research draws the assumption of a hierarchy between the linguistic parameters used in signs as meaningful elements.

If those greatly iconic structures involve infra-lexical linguistic elements that do not belong to the lexicon, they appear most often in narrating sequences and remain nonetheless unmentioned in Sign Language dictionaries. However, if we consider that these structures are an integral part of Sign Language, how should we integrate them into our web site?

Two perspectives are suggested to answer this question. The first one consists in considering the minimal structures of realization in Sign Language. The linguistic parameters of configuration, movement, location (Stokoe et al., 2000) and orientation (Friedman, 1977), (Liddell, 1980), (Moody, 1983), (Yau, 1992) cannot be considered as such. Indeed, even if a human mind can make a relevant distinction between them as isolated elements conveying meaning, they must be used simultaneously in order to be activated while communicating. Contrary to vocal languages, realizing a signifying form in a Sign Language cannot be made through a succession of distinct realizations of isolated and non-signifying elements. Minimal realization structures in Sign Language may be ranged on a growing complexity scale, starting from the formal transfer (infra-conceptual level) and going up to the double transfer (level where several actors, location parameters and utterances can be combined). These various structures use the same linguistic parameters during the same realization laps of time.

The second perspective is based on the dialectics between syntagm and paradigm. When narratives contain highly iconic structures, the value of an element at a given time undergoes a type of syntagmatic pressure – which does not necessarily come from preceding or following units, but from other units occurring at the same time and taking part in the minimal form of realization as well. Yet, the simultaneity is not a sufficient clue to conclude that it is a paradigm, since this pressure can be seen. In a Sign Language, the pressure stemming from the context does not only influence the temporal dimension. The spatial dimension exerts constraints as well, but this time simultaneous instead of successive. Regarding these two perspectives, our users are questioned about their perceptions and representations of the meaningful infra-conceptual units – while first visiting each "sign page".

We do not want to collect "correct" answers, but to gather the most identical ones. This way, our classification leans on a consensus amongst users. However, our experts can impose a classification and may concentrate the researches for a sign through these answers, without necessarily using the material as a final classification.

Based on our users' answers, descriptors and/or classifiers are assigned to each sign, according to the summary of a dynamic amount of identical and meaningful answers. The data base model we propose is based on the idea of modelling the interactions giving sense to the content – and not the content itself. The polymorphic use of contents implies a data organization based on the role we wish a content to play, as explained above. In this way, an "answer" – as a content – has both roles of answer to a question and of classifying and research element. One or several specific descriptors correspond to each role.

This approach of a dynamic classification of concepts, built upon LSF linguistic parameters specific to each sign, enables us to propose our users to look up concepts through the site directly in LSF, without having to know the concepts' written signifiers. Later on, a dynamic classification could also be based upon sign writing[4].

4. References

A. Bénel. 2003. *Consultation assistée par ordinateur de la documentation en Sciences Humaines : Considérations épistémologiques, solutions opératoires et applications l'archéologie.* Ph.D. thesis. Thèse de doctorat en informatique, Institut National des Sciences Appliquées de Lyon, 2003.

J. Bideaud, O. Houde, and J.-L. Pedinielli. 1993. *L'homme en développement.* PUF, Paris.

C. Courtin. 1998. *Surdité, langue des signes et développement cognitif.* Ph.D. thesis, Université Paris V.

C. Cuxac. 2000. *La Langue des Signes Francaise : les voies de l'iconicité.* OPHRY, PARIS.

P. Dalle. 2003. La place de langue des signes dans le milieu institutionnel de léducation: enjeux, blocage et évolution. *Langue franaise*, 137.

F. Duquesne-Belfais. 2007. *Activité et langages dans la conceptualisation mathématique : des apprentissages des élèves sourds la formation de leurs enseignants.* Ph.D. thesis, Université Lille 1.

Lynn A. Friedman. 1977. *On the other hand : new perspectives on American sign language.* Academic Press, New York.

S.K. Liddell. 1980. *American Sign Language Syntax.* The Hague.

B. Moody. 1983. *Histoire et Grammaire, tome I.* Ellipses, Paris.

C. Moreau and B. Mascret. 2008. Lexique LSF. In Onno et al (eds) Crasborn, editor, *3rd Workshop on the Representation and Processing of Sign Languages. 6th International Conference on Language Resources and Evaluation, LREC 2008*, Construction and Exploitation of Sign Language Corpora., pages 138–140, June.

C. Moreau. 2008. Lexiquelsf : vers une web académie de la langue des signes franaise. *La nouvelle revue de l'adaptation et de la scolarisation*, 43, October.

W.C. Stokoe, D.C. Casterline, and C.G. Croneberg. 2000. In *A dictionary of american sign language*, WASHINGTON D.C. Gallaudet College Press.

S.-C. Yau. 1992. *Création Gestuelle et début du Langage Création de langues gestuelles chez les sourds isolés.* Éditions Langages Croisés, Hong Kong.

---

[4]http://www.signwriting.org/