What we can learn from exploring health-related texts : the example of speech and language therapy (SLT)

Frederique Brin-Henry, SLT PhD

frederique.henry@atilf.fr

From translation and lexicography to

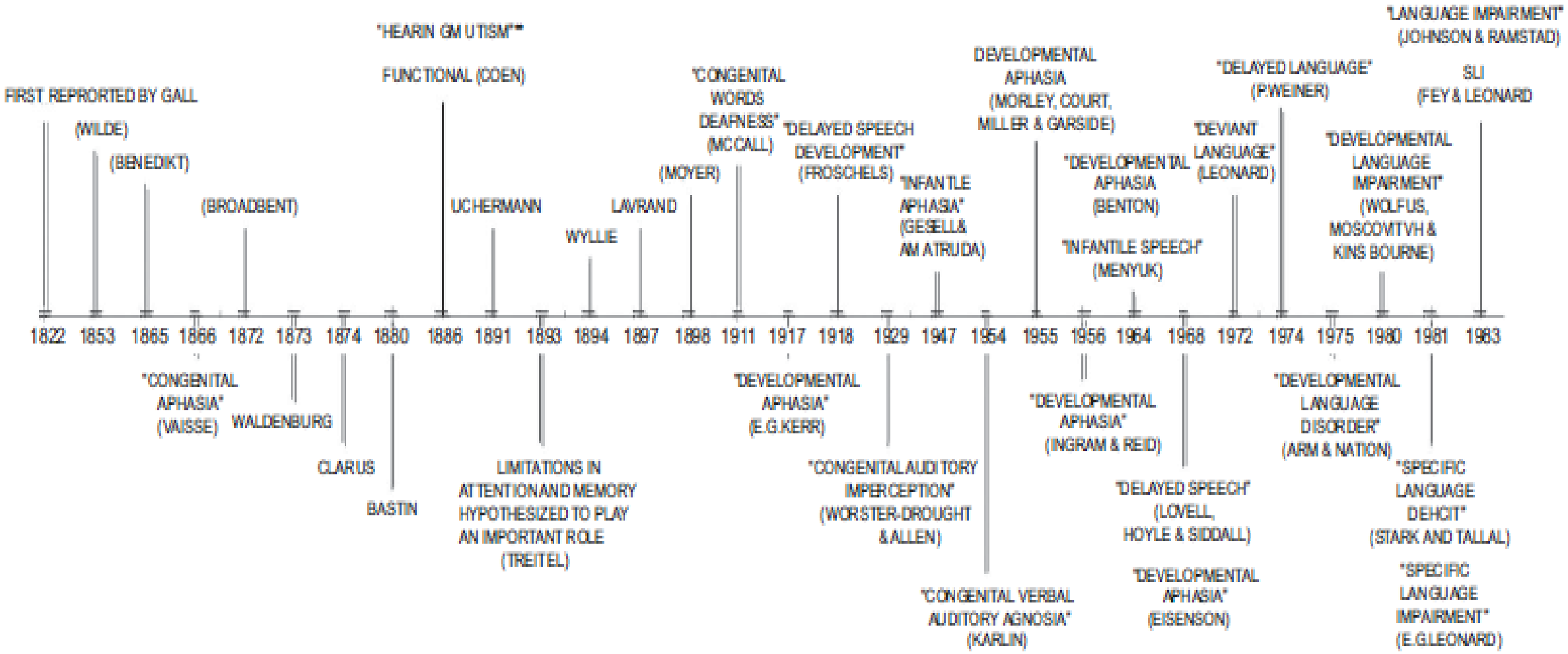
Terminology
Terminography

What speech and language therapists say, and how they talk about:

- language pathology
- their patients and their professional activities
- their discipline

Termes génériques	Termes relatifs aux troubles de la lecture et de l'orthographe	Formulations contenant des relations causales ou comorbidité	Termes étiquetant des sous-types de troubles de la lecture et de l'orthographe
Learning difficulties (Poland) **Literacy impairment** (USA) **Poor academic skills** (India) **Specific academic needs** (Bulgaria) **Specific impairments of academic skills** (Bulgaria) **Specific Learning Difficulties** (Malta) **Specific Learning Disabilities** (Slovakia) **Specific learning disabilities in reading/writing** (Australia) **Specific learning Impairments** (Bulgaria)	**Disorder of written expression: DSM-IV criteria** (Canada) **Dysgraphia-dyslexia** (Bulgaria) **Legasthenie** (Austria, Switzerland) **Reading and writing disorders** (Slovenia) **Reading disorder** (Bulgaria, Canada, Latvia) **Reading retarded** (Denmark) **Specific disorders of the written language** (France) **Specific reading disabilities** (Latvia) **Word blind** (Denmark) **Written language disorder** (Australia) **Written language Impairment** (Malta)	**Language based reading disability** (Canada) **Language learning disabilities** (USA) **Oral and written language impairment/disability** (USA) **Reading and writing disabilities due to insufficient development of the language system** (Latvia) **specific language impairments** (Sweden)	**Phonological dysgraphia** (Russia) **Phonological subtypes** (Germany) **Regulatory dysgraphia** (Russia) **Visual-spatial dysgraphia** (Russia)

Table 2: termes en usage concernant les troubles du langage écrit (Issafova 2013)



Reilly S., Tombin B., Law J., McKean C., Mensah F.K., Morgan A., Goldfeld S., Nicholson J.M., Wake M. (2014) Specific language impairment: a convenient label for whom? *Int J Lang Commun Disord*, *49* (4), 416–451

Classifications used in French SLT

Dyslexie (DSM-5) : trouble spécifique des apprentissages avec déficit en lecture = *dyslexie*
(vs Trouble non spécifique des apprentissages avec déficit de la lecture = *retard en lecture* (Launay, 2018)

R48.0 : dyslexie et alexie (CSARR)

NGAP : trouble de la communication et du langage écrit

Synonyme CISMeF : trouble d'acquisition de la lecture; Alexie verbale; difficulté d'apprentissage de la lecture;

Synonyme MeSH : Légasthénie; Cécité verbale pure; Trouble de la lecture; Dyslexie développementale; Trouble de l'apprentissage de la lecture;

Hyponyme MeSH : Trouble de l'acquisition de la lecture; Alexie;

Terminology

The central object of terminology are :

- terminological units (Cabré 2003),
- or concepts materialized by linguistic labels (Felber 1980 ; Wüster 1974).

Concept:

- unit of knowledge created by a unique combination of characteristics

Term

- entity with a form and a meaning
- ISO 1087: Verbal designation of a general concept in a specific subject field

Methods

What to work on? SLT practice (assessment reports) + published specialised documents

The challenge of building corpora

Discourse analysis/ textual analysis/ extracting and studyoing occurrences on a semantic, syntactic and terminological level

Link between expert opinion/validation and automatic extraction

Studying SLT Terminology

Towards a conceptual representation of language pathology based on properties extracted from SLT specialty language

KEY ISSUES

- At a clinical level

SLT diagnostic shows specific features

- Importance of assessment
- Deliberate mention of potentialities
- Variation of terminology
- Importance of temporality (verbs), durability, onset time, evolution and synchronicity

- At a scientific level

SLTs terminology hold specific linguistic characteristics

- Frequency of words outside classifications (ie *difficultés*)
- Frequency of some syntagms: N de N de N
- Clinical nuances outside linguistic relevance : *trouble d'articulation vs trouble de l'articulation*

Resources

Lexical and terminological resources
OrthoCorpus

Termino-ontology
TemPO

General information on the research work

Observatoire terminologique en orthophonie OTO

https://www.ortolang.fr/market/corpora/orthocorpus/v2

https://oto-fr.atilf.fr/

Applications

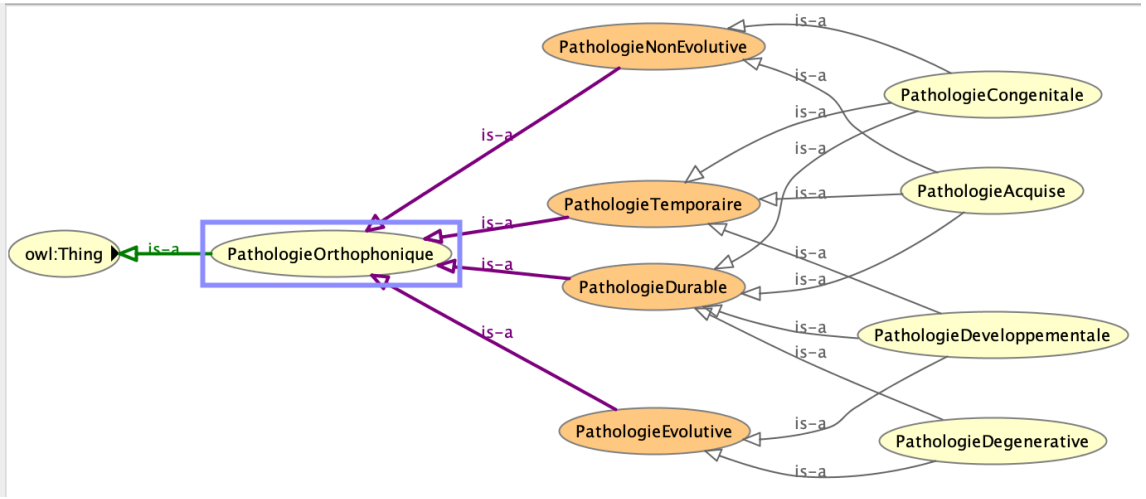
- At a clinical level

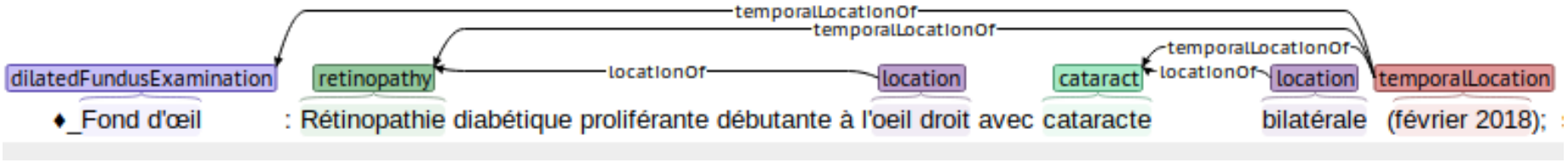
- Improve communication between allied health professionals
- Improve clinical research on available data (making sure we are talking about the same thing)

Applications

- At a clinical level
- At a scientific level

- Improve classifications and description of language pathology in French and across languages
- Improve access to health related documents





The CLEAR project

CLEAR (Communication, Literacy, Education, Accessibility, Readability)

ANR-17-CE19-0016-01.

CLEAR

innovative methods allowing creation of linguistic resources and software dedicated to the simplification of medical texts written in French

- Conducted by: N. Grabar CR1 CNRS, UMR 8163 STL
  http://natalia.grabar.free.fr/
- Stakeholders :

  AFH ASSOCIATION DES HEMOPHILES
  SYNAPSE SYNAPSE DEVELOPPEMENT
  LIMSI Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
  LEPS EA 3412 LABORATOIRE EDUCATIONS ET PRATIQUES EN SANTÉ
  MESHS - STL UMR8163 Maison Européenne des Sciences de l'Homme et de la Société Lille Nord-de-France/STL-Savoirs, Textes, Langage

CLEAR

Methods for creating resources and a prototype

Challenges :

- Focusing on patients needs for information
- Using large quantities of heterogenous non-structured data,
- Using NLP methods for the medical field
- Creating a Knowledge base for explicit medical terms in French.

Goals :

- improve interactions between patient and health professionals,
- Facilitate and increase Health literacy

CLEAR

- Identification of complex segments and sentences
- Provide lexicons for alternative options
- Automatic production of readable texts
- Accessible prototype

CLEAR : methods

- Creation of a corpus : 10 000 couples of aligned sentences
- Researching parallel sentences in comparable texts (using lexical descriptors and chain of characters)
- Detection of difficult segments and words
- Creation of a lexicon linking technical medical terms with lay equivalents
- Automatic simplification of medical documents (+ pictures)

The POPEX project
Extraction d'informations sur les populations

POPEX

- At a clinical level
  - Formalising use cases
  - Facilitating collection of data within non-structured texts (quality of care, professional practice analysis)
- At a scientific level
  - Principle :Identification of medical entities (+ relations And events) within documents
  - Results : information available in a structured format
  - Information extracted according to the user needs/interests

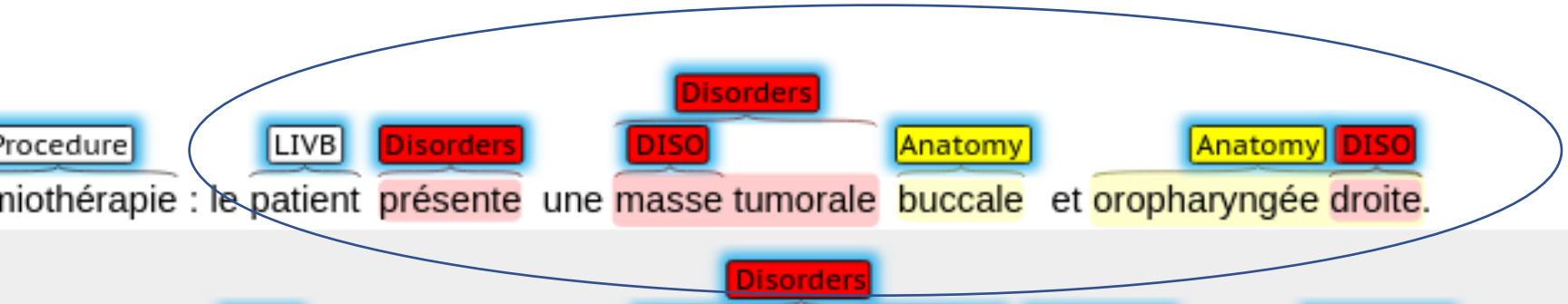
POPEX

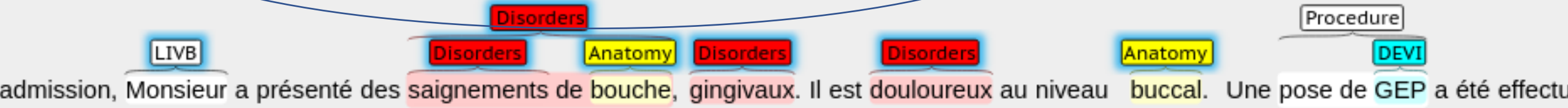
- Main challenges :
  - Access to de-identified texts (GDRP General Data Protection Regulation)
  - Quality/ nature of segments (SLT is a subdomain a specialized medical domain) :
    - Which named entities
    - Which lexicon
    - Size of lexicon for deep learning (QuickMatching + CamemBERT model trained with QUAERO + other corpora)
  - Discussion on role of health professionals/automatic procedures

Medical and SLT corpora

Corpus	Description	Taille
OrthoCorpus (2019)	Articles de la revue spécialisée Rééducation Orthophonique (Brin-Henry, 2018)	6,7M
ISTEX	Articles de revues médicales indexées par ISTEX (Inist)	42,6M
EQueR	Articles scientifiques et de recommandations de bonne pratique médicale (CISMeF)	16,8M
PMC OA	Articles de revues médicales (PubMed Central Open Access)	3,8M
Cochrane	Résumés d'articles de l'organisation Cochrane	5,0M
EMA	Notices de l'Agence Européenne des Médicaments	21,2M
CRTT	Articles de revues, extraits de Science Direct	21,7M
E3C-Corpus	Résumés d'articles, articles de revues, cas cliniques	12,1M
Wikipédia	Articles Wikipédia dans le domaine médical	6,6M

TABLE 1 – Collections de textes médicaux utilisées. Les tailles sont exprimées en millions de mots


Procedure
LIVB
Disorders
Disorders
DISO
Anatomy
Anatomy
DISO
niothérapie : le patient présente une masse tumorale buccale et oropharyngée droite.


Disorders
LIVB
Disorders
Anatomy
Disorders
Disorders
Anatomy
Procedure
DEVI
admission, Monsieur a présenté des saignements de bouche, gingivaux. Il est douloureux au niveau buccal. Une pose de GEP a été effectu

2 Le dossier documente plusieurs antécédents (éthylotabagisme, cardiopathie, HTA, AVC droit survenu en XXXX). Depuis son admission, Monsieur a présenté des saignem

Disorders Physiology Physiology PROC Disorders
des difficultés de mastication, mais l'alimentation per os est maintenue en texture molle, il n'y pas de fausses routes.

LIVB Procedure Disorders
3 Le patient se montre très coopérant durant l'entretien, on ne relève aucune fatigabilité.

Physiology Physiology
5 Bilan de la déglutition et des fonctions oromyofonctionnelles :

Disorders Disorders Anatomy Anatomy LIVB Disorders Disorders Anatomy Disorders
7 La masse tumorale est volumineuse au côté droit et encombre la zone intra buccale, le patient présente une asymétrie labiale. D'autre part, on observe une fente palatine

Procedure Physiology Procedure Signe Signe
PROC Anatomy Anatomy
8 - Praxies bucco faciales : déficit sur le plan de la motricité linguale avec une protrusion très déficitaire, une latéralisation à gauche difficile. L'élévation et l'abaissement sont

Procedure PROC
DISO
9 - La toux sur demande est tonique et semble efficace. L'apnée est possible.

Disorders
Physiology Procedure Signe
10 - Articulation : les nasales sont légèrement oralisées. La réalisation du [k] est satisfaisante. Intelligibilité réduite en contexte conversationnel (voir compte-rendu de bilan de l

PHYS Disorders Anatomy
11 - Voix : intensité, fréquence normales. Timbre altéré par une anomalie de la résonance ; aspect « engorgé » de la parole.

Physiology Anatomy Physiology Devices
12 - déglutition automatique et sur commande observées (salivaire): élévation laryngée et contraction des masséters satisfaisants.

t peu dans les contacts mais de façon pertinente. Les informations fournies, nombreuses, par oral et à l'écrit, n'ont peut-être pas été mémorisées (non vérifié). Elle communique principalement par gestes, ou par

mais elle a du mal à en sélectionner de nouveaux ou à concentrer sa demande. Aucune aide technique à la communication n'a été envisagée.

c une difficulté à obtenir une fermeture labiale (syncinésie œil), on obtient une petite protrusion linguale, l'élévation linguale est minime (seul l'apex).

ée et soutenu, on entend des stases salivaires assez importantes et régulières. La patiente déglutit la salive sur ordre et de façon automatique. L'articulation des phonèmes oraux isolés est impossible, on

ant une hyperrhinophonie lors de l'émission des voyelles. Aucune consonne n'est produite sur ordre en raison de la difficulté à produire l'occlusion ou la constriction nécessaire.

s routes fréquentes aux liquides. Elle a spontanément adapté son alimentation mais conserve des aliments à risque. Un essai à l'eau gélifiée montre une évolution difficile du bol alimentaire en bouche, avec

e descend pas et on note des stases pharyngées importantes. La toux se déclenche mais n'est pas efficace ; la patiente met plusieurs secondes à reprendre son souffle. L'alimentation per os requiert donc une

Thank you !

Frederique Brin-Henry, SLT PhD

frederique.henry@atilf.fr

references

- BISHOP, D. V., SNOWLING, M., THOMPSON, P. A., & GREENHALGH, T. 2017. Phase 2 of CATALISE : a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J Child Psychol Psychiatry, 58*(10):1068-1080.
- BRIN-HENRY F. (2018). Pour une harmonisation de la terminologie orthophonique : contribution du projet OrthoCorpus (2015- 2017). In Terminologica. TOTh 2018.
- BRIN-HENRY F., COSTA R., DESPRES S. TemPO : towards a conceptualisation of pathology in speech and language therapy. *Terminologie et Intelligence Artificielle – Atelier TALN-RECITAL et IC (PFIA 2019)*, Jul 2019, Toulouse, France. ⟨hal-02193859⟩
- COSTA, R. (2013). Terminology and specialized lexicography : two complementary domains. *Lexicographica, 29* (1), 29-42.
- LINDBERG D. A., HUMPHREYS B. L. & MCCRAY A. T. (1993). The unified medical language system. Methods of information in medicine, 32(4), 281–291.
- NEVEOL A., GROUIN C., LEIXA J., ROSSET S. & ZWEIGENBAUM P. (2014). The QUAERO French medical corpus : A ressource for medical entity recognition and normalization. In Proc of BioTextMining Work, p. 24–30.