

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**

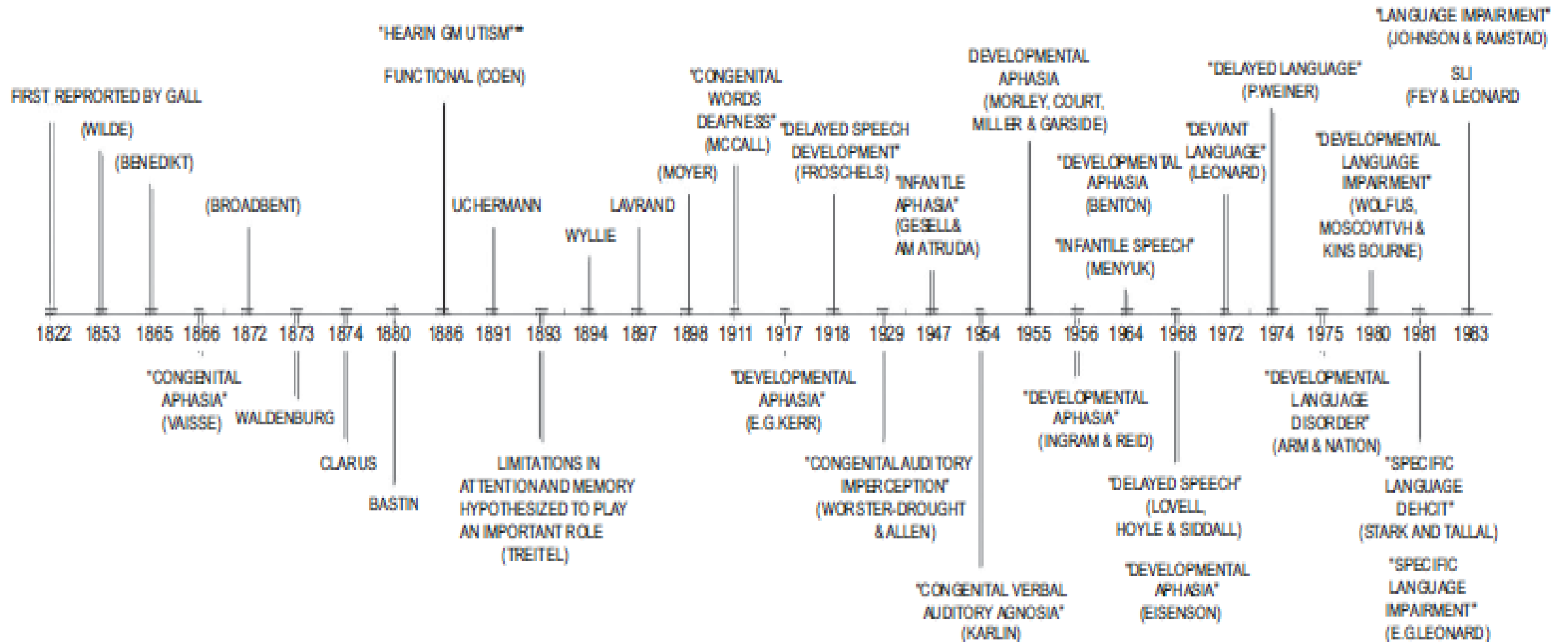
**Terminology
Terminography**

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

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

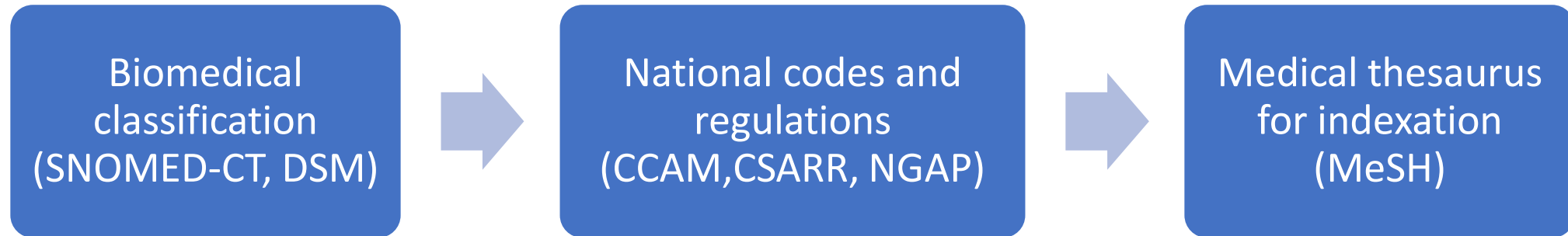
Termes génériques	Termes relatifs aux troubles de la lecture et de l'orthographe	Formulations contenant des relations causales ou comorbidité	Termes étiquetant des sous-types de troubles de la lecture et de l'orthographe
Learning difficulties (Poland)	Disorder of written expression: DSM-IV criteria (Canada)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

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 (Cabr  2003),
- or concepts materialized by linguistic labels (Felber 1980 ; W ster 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

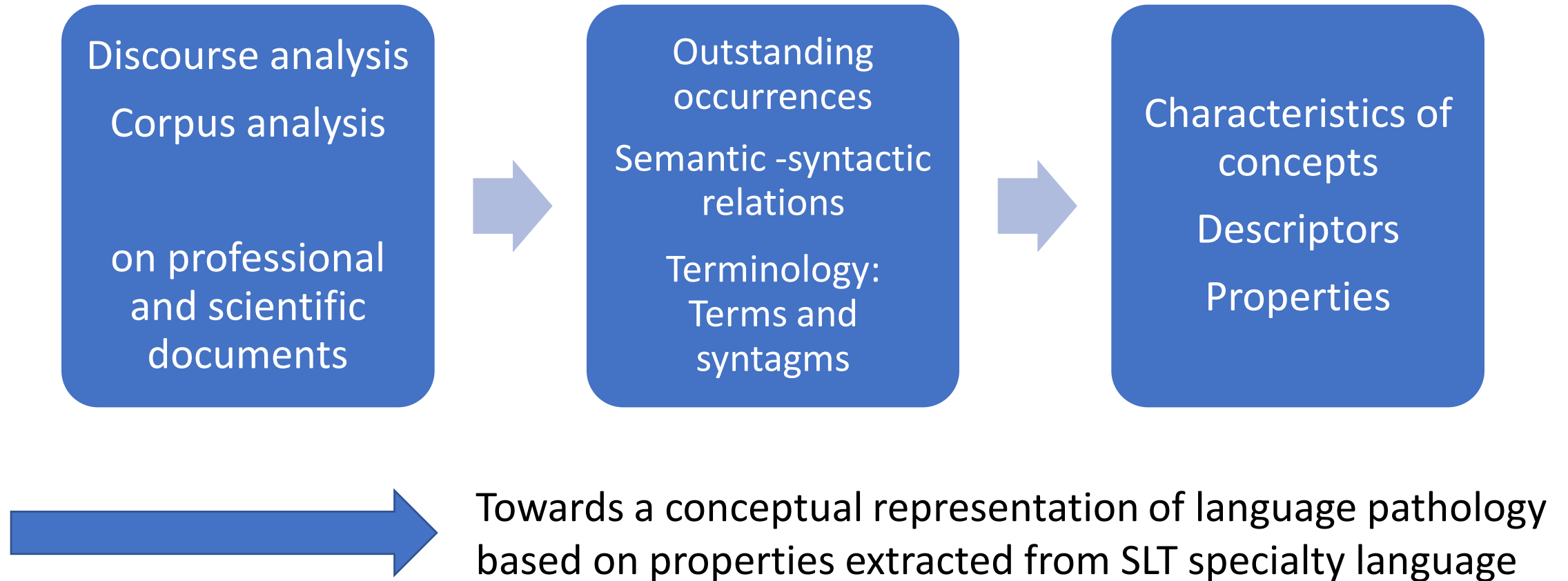


Discourse analysis/ textual analysis/
extracting and studying occurrences on a
semantic, syntactic and terminological level



Link between expert opinion/validation and
automatic extraction

Studying SLT Terminology



KEY ISSUES

- At a clinical level
- At a scientific 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

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

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

Termino-ontology
TemPO

General information on
the research work


Observatoire
terminologique en
orthophonie OTO

<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)



Dictionnaire d'Orthophonie
Prédérique BÉGIN-HENRI • Catherine COURRIER • Emmanuelle LECORLE • Véronique MASY


Mon compte Se déconnecter À propos Mode d'emploi

 **Rechercher**

Mot du jour
Contraintes intralinguistiques

Définition
Contraintes relatives à la langue et qui permettent une communication efficace. Le nombre restreint d'unités linguistiques (phonèmes, morphèmes, syntagmes) et des règles combinatoires qui régissent leur intégration dans la chaîne parlée contraignent le locuteur à réaliser des productions verbales (sonnées ou fortes) beaucoup moins nombreuses que ce qu'il pourrait virtuellement exprimer. Le souci de mettre du sens dans ce qu'on exprime apporte également des restrictions à nos énoncés possibles : le chat attrape le lapin ; ?). Jakobson a montré que plus les unités linguistiques sont « petites », plus la contrainte est forte, le choix et la combinaison des traits articulatoires pour la production des phonèmes offrant peu de liberté au locuteur, contrairement à ce qui se passe pour les éléments syntaxiques et surtout narratifs.

En savoir plus →



GERONTOPOLE
SERVICE DE GERIATRIE INTERNE
201 AVENUE DE TANGUET - TSA 4001
31050 TOULOUSE
Tél : 05 61 12 14 15
www.hopital-toulouse.fr

DR PHILIPPE PUYBARET
60 AVENUE JEAN CHAUBET
31500 TOULOUSE

Toulouse, le 06/08/2014

Références : OTO/CEY

Cher Confrère,

Votre patient **Monsieur Rogelio VAZQUEZ**, né le 14/04/1928 (86 ans), a été hospitalisé dans le Service de Gériatrie Clinique, secteur D, du 4/08/2014 au 5/08/2014, dans les suites d'une décompensation cardiaque.

Mode de vie :
- Monsieur V. vit à domicile avec sa compagne.

Monsieur Rogelio VAZQUEZ présente dans ses antécédents :
- une cardiopathie malade ischémique et valvulaire (insuffisance mitrale et tricuspidienne),
- une fibrose pulmonaire à la CORDARONE,
- une fibrillation auriculaire permanente.

Histoire de la maladie :
- Monsieur V. a été hospitalisé du 2 au 4/08/2014 dans le service des soins intensifs de cardiologie de l'hôpital de Rangueil. Le diagnostic de décompensation cardiaque droite est réalisé. L'étiologie semble être une rupture du régime sans sel lors d'un séjour en Espagne (prise de 7 kg). La mise en place de diurétiques par voie intraveineuse continue, a permis une amélioration des symptômes autorisant son transfert dans notre unité.

A l'entrée dans le service :
L'examen clinique à l'entrée retrouve :

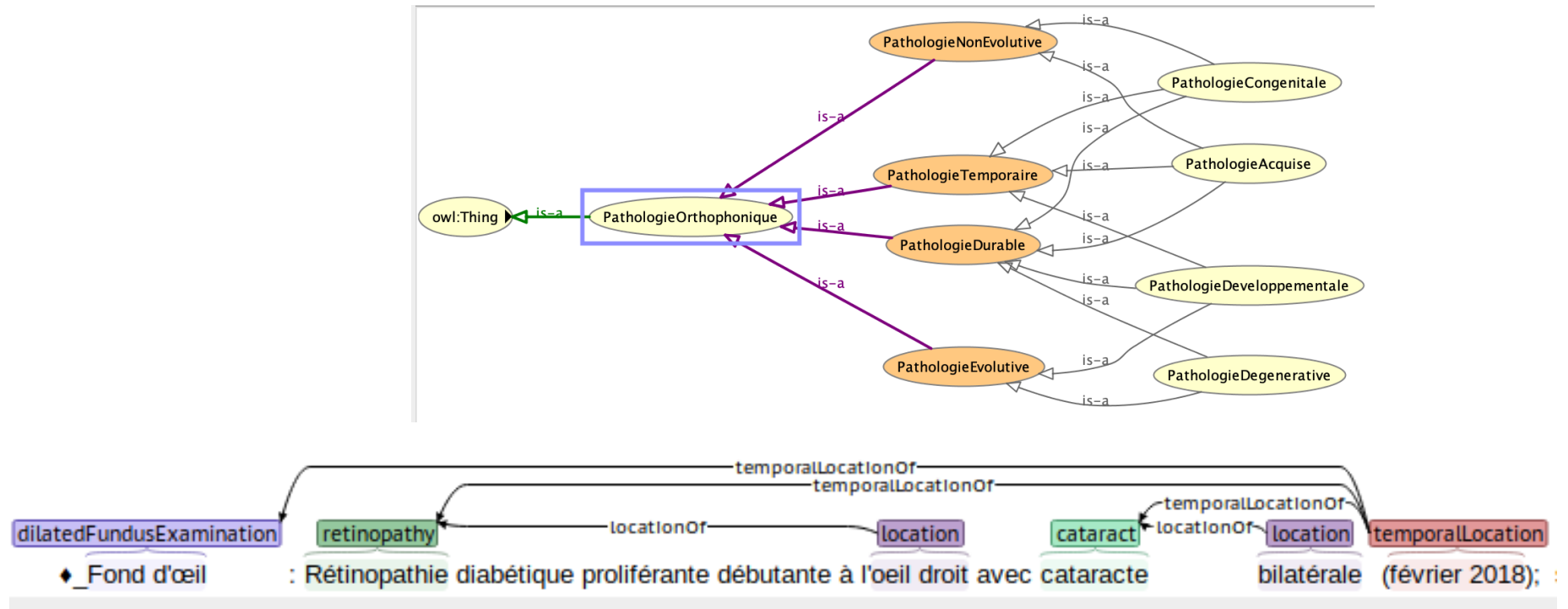
- * Sur le plan neurologique :
L'examen retrouve un patient conscient, bien orienté. Il n'y a pas de déficit sensitivo moteur, pas de syndrome extrapyramidal.
- * Sur le plan cardio-vasculaire :
Les bruits du cœur sont irréguliers, à 80/min avec un souffle systolique. On note la présence d'importantes turgescences jugulaires bilatérales. Il présente également de discrets œdèmes des membres inférieurs associés à une ascite modérée.
- * Sur le plan respiratoire :
L'auscultation pulmonaire retrouve des crépitations au niveau des bases avec une discrète dyspnée de repos. La tension artérielle est à 12/7 avec une saturation à 98 % en air ambiant.

Validé électroniquement par : Dr TOULZA Olivier le 06/08/2014
Consultez le site internet du Centre Hospitalier Universitaire : www.chu-toulouse.fr

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

un **syndrome bulbaire** avec **atteintes des paires crâniennes**. La patiente **présente** un **faciès** peu expressif, **bouche** ouverte,



POPEX

- Main challenges :
 - Access to de-identified texts (GDPR 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 DISO Anatomy Anatomy DISO
niothérapie : le patient présente une masse tumorale buccale et oropharyngée droite.

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 effectuée.

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

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

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

5 Bilan de la déglutition et des fonctions oromyofonctionnelles :

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

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

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

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

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

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

PROC

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

Devices Devices

Procedure

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.

Physiology

Anatomy

Disorders

ANAT

DISO

Physiology

Anatomy

Procedure

Anatomy

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

Disorders

DISO

Anatomy

LIVB

Physiology

PHYS

ée et soutenue, 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

Anatomy

Phenomenon

Phenomenon

Disorders

Physiology

Anatomy

Anatomy

Physiology

Physiology

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.

Disorders

CHEM

Objects

Objects

PROC

Objects

Physiology

Anatomy

Anatomy

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

Disorders

DISO

LIVB

PHYS

Procedure

ANAT

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 resource for medical entity recognition and normalization. In Proc of BioTextMining Work, p. 24–30.