


UDC: 7203.01

LBC: 63.3(2)6-7; 65.497

MJ № 479

 10.33864/2617-751X.2026.v9.i2.381-398

## THE IMPORTANCE OF CORPUS LINGUISTICS IN SPEECH AND LANGUAGE THERAPY PRACTICE IN ALGERIA

Mouloud Hadbi\*

**Abstract.** This article explores the importance of corpus linguistics in the training and professional practice of speech and language therapists (SLTs) in Algeria. It highlights how the integration of corpus-based methods can enhance both academic learning and clinical efficiency. Through a historical and methodological overview, the paper emphasizes that corpus linguistics provides a scientific framework for observing, analyzing, and interpreting authentic language data. In the Algerian multilingual context - where Arabic, Berber, and French coexist - corpus approaches enable practitioners to better differentiate between linguistic variation and actual language disorders. The article also examines the pedagogical benefits of including corpus linguistics in SLT curricula, notably in developing critical thinking, reflective practice, and data-handling skills. By promoting the systematic use of linguistic corpora in diagnosis, therapy, and research, this approach contributes to the emergence of a more contextualized and evidence-based speech therapy practice in Algeria. The conclusion calls for stronger institutional support and training initiatives to build a national corpus culture that reflects the linguistic diversity of the country.

**Keywords:** Corpus linguistics; Language disorders; Multilingualism; Reflective training

---

\* University of Abderrahmane Mira, Bejaia; Algeria

E-mail: [mouloud.hadbi@univ-bejaia.dz](mailto:mouloud.hadbi@univ-bejaia.dz)

<https://orcid.org/0009-0000-8119-8407>

**To cite this article:** Hadbi, M. [2026]. THE IMPORTANCE OF CORPUS LINGUISTICS IN SPEECH AND LANGUAGE THERAPY PRACTICE IN ALGERIA. "Metafizika" journal, 9(2), pp.381-398.

<https://doi.org/10.33864/2617-751X.2026.v9.i2.381-398>

**Article history:**

**Received:** 09.06.2025 **Revised:** 19.10.2025 **Accepted:** 06.12.2025 **Published:** 15.06.2026




**Copyright:** © 2025 by AcademyGate Publishing. This article is an open access article distributed under the terms and conditions of the CC BY-NC 4.0. For details on this license, please visit

<https://creativecommons.org/licenses/by-nc/4.0/>.

УДК: 7203.01

ББК: 63.3(2)6-7; 65.497

МЖ № 479

 10.33864/2617-751X.2026.v9.i2.381-398

## ЗНАЧЕНИЕ КОРПУСНОЙ ЛИНГВИСТИКИ В ПРАКТИКЕ ЛОГОПЕДИИ И РЕЧЕВОЙ ТЕРАПИИ В АЛЖИРЕ

Мулуд Хадби\*

**Абстракт.** В статье рассматривается значение корпусной лингвистики в профессиональной подготовке и практической деятельности логопедов и специалистов по речевой терапии в Алжире. Показано, что интеграция корпусно-ориентированных методов способствует повышению как академического уровня обучения, так и эффективности клинической практики. На основе историко-методологического обзора подчеркивается, что корпусная лингвистика предоставляет научно обоснованную платформу для наблюдения, анализа и интерпретации аутентичных языковых данных. В условиях алжирского многоязычия - сосуществования арабского, берберского и французского языков - корпусные подходы позволяют более точно различать языковую вариативность и собственно речевые нарушения. В статье также анализируются педагогические преимущества включения корпусной лингвистики в образовательные программы по логопедии, в частности развитие критического мышления, рефлексивной практики и навыков работы с данными. Продвижение систематического использования лингвистических корпусов в диагностике, терапии и научных исследованиях способствует формированию контекстуализированной и доказательной модели логопедической практики в Алжире. В заключении подчеркивается необходимость усиления институциональной поддержки и развития программ профессиональной подготовки, направленных на формирование национальной корпусной культуры, отражающей языковое разнообразие страны.

**Ключевые слова:** корпусная лингвистика; речевые нарушения; многоязычие; рефлексивная подготовка

---

\* Университет Абдеррахмана Мира, Беджая; Алжир

E-mail: [mouloud.hadbi@univ-bejaia.dz](mailto:mouloud.hadbi@univ-bejaia.dz)

<https://orcid.org/0009-0000-8119-8407>

**Цитировать статью:** Мулуд, Х. [2026]. ЗНАЧЕНИЕ КОРПУСНОЙ ЛИНГВИСТИКИ В ПРАКТИКЕ ЛОГОПЕДИИ И РЕЧЕВОЙ ТЕРАПИИ В АЛЖИРЕ. Журнал «Metafizika», 9(2), с.381-398.

<https://doi.org/10.33864/2617-751X.2026.v9.i2.381-398>

**История статьи:**

Поступила: 09.06.2025 Переработана: 19.10.2025 Принята: 06.12.2025 Опубликовано: 15.06.2026




Copyright: © 2025 by AcademyGate Publishing. This article is an open access article distributed under the terms and conditions of the CC BY-NC 4.0. For details on this license, please visit

<https://creativecommons.org/licenses/by-nc/4.0/>.

UOT: 7203.01

KBT: 63.3(2)6-7; 65.497

MJ № 479

 10.33864/2617-751X.2026.v9.i2.381-398

## ƏLCƏZAİRDƏ NİTQ VƏ DİL TERAPİYASI PRAKTİKASINDA KORPUS LİŦVİSTİKASININ ƏHƏMİYYƏTİ

Mulud Hadbi\*

**Abstrakt.** Məqalədə Əlcəzairdə nitq və dil terapevtlərinin (loqopedlərin) peşəkar hazırlığında və praktik fəaliyyətində korpus linqvistikasının əhəmiyyəti araşdırılır. Korpus əsaslı metodların tətbiqinin həm akademik təlimi, həm də klinik fəaliyyətin səmərəliliyini artırdığı göstərilir. Tarixi və metodoloji icmal əsasında vurğulanır ki, korpus linqvistikası autentik dil materiallarının müşahidəsi, təhlili və şərhini üçün elmi əsaslandırılmış çərçivə təqdim edir. Əlcəzairin çoxdilli mühitində - ərəb, bərbər və fransız dillərinin yanaşı mövcud olduğu şəraitdə - korpus yanaşmaları dil variasiyasını real nitq və dil pozuntularından daha dəqiq ayırmağa imkan verir. Məqalədə həmçinin korpus linqvistikasının loqopediya üzrə tədris proqramlarına daxil edilməsinin pedaqoji üstünlükləri, xüsusilə tənqidi düşünmə, reflektiv praktika və məlumatlarla işləmə bacarıqlarının inkişafı təhlil olunur. Diaqnostika, terapiya və elmi tədqiqatlarda linqvistik korpuslardan sistemli istifadə Əlcəzairdə daha kontekstual və sübutlara əsaslanan nitq terapiyası praktikasının formalaşmasına töhfə verir. Nəticə hissəsində ölkənin dil müxtəlifliyini əks etdirən milli korpus mədəniyyətinin yaradılması üçün institusional dəstəyin gücləndirilməsinin və ixtisasartırma təşəbbüslərinin vacibliyi qeyd olunur.

**Açar sözlər:** korpus linqvistikası; dil pozuntuları; çoxdillilik; reflektiv təlim

---

\* Abderrahmane Mira Universiteti, Becaia; Əlcəzair

E-mail: [mouloud.hadbi@univ-bejaia.dz](mailto:mouloud.hadbi@univ-bejaia.dz)

<https://orcid.org/0009-0000-8119-8407>

**Məqaləyə istinad:** Hadbi., M. [2026]. ƏLCƏZAİRDƏ NİTQ VƏ DİL TERAPİYASI PRAKTİKASINDA KORPUS LİŦVİSTİKASININ ƏHƏMİYYƏTİ. “Metafizika” jurnali, 9(2), səh.381-398.

<https://doi.org/10.33864/2617-751X.2026.v9.i2.381-398>

**Məqalənin tarixçəsi:**

**Daxil olub:** 09.06.2025 **Yenidən baxılıb:** 19.10.2025 **Təsdiqlənib:** 06.12.2025 **Dərc olunub:** 15.06.2026



**Copyright:** © 2025 by AcademyGate Publishing. This article is an open access article distributed under the terms and conditions of the CC BY-NC 4.0. For details on this license, please visit

<https://creativecommons.org/licenses/by-nc/4.0/>.

## 1. Introduction

Corpus linguistics in speech therapy analyzes linguistic data (both oral and written) collected during therapeutic follow-ups in order to understand and treat language disorders. It relies on personalized and authentic corpora to identify linguistic variations associated with a specific pathology, thereby contributing to a better understanding and assessment of linguistic deviations within a clinical context. The use of corpus linguistics allows for the detailed analysis of language productions, facilitating an in-depth study of patients' linguistic performance, including syntax, semantics, and pragmatics. These corpora can be compiled from data gathered during therapeutic sessions.

It also facilitates the personalization of assessment and rehabilitation, as the analysis of individualized corpora helps capture the specificities of each patient's speech, providing a foundation for more tailored evaluation and intervention strategies. Moreover, it enables speech therapists to gain a deeper understanding of pathologies by comparing patients' linguistic data with reference corpora, thereby identifying the status of the pathology and the linguistic deviations associated with it.

Corpus linguistics also optimizes therapeutic intervention, as it allows clinicians to question the relevance of remediation methods based on the linguistic features identified, ultimately aiming to enhance the effectiveness of therapeutic practices.

In speech-language pathology, two main types of corpora can be distinguished: reference corpora and therapeutic corpora. The former consist of a wide variety of texts representing diverse uses of the language. They serve as benchmarks for understanding the general functioning of a language and can be used to compare with patient corpora. The latter, in contrast, are created specifically from data collected from patients during their speech therapy sessions.

In this article, we will focus on the training of Algerian students in corpus linguistics - whether it is sufficient - and on the contextualization of this tool within speech therapy practice in Algeria. To what extent do speech therapists apply this tool in their daily practice?

Before addressing these questions, we will begin by defining what a corpus is and providing a brief historical overview of this approach.

### 1) Definition of Corpus Linguistics

According to *Le Trésor de la langue française*, the main reference dictionary of the French language, a corpus in the field of linguistics is defined as “a collection of texts established according to a principle of exhaustive documentation, a thematic or exemplary criterion, with a view to their linguistic study” [<https://www.le-tresor-de-la-langue.fr/definition/corpus#top>, consulted on 04/10/25 at 19:00].

In corpus linguistics, the most frequently cited definition among specialists is that of John Sinclair, an influential British linguist, who describes a corpus as “a collection of language data selected and organized according to explicit linguistic criteria to serve as samples of language use.” This linguistic variety forms the basis of computer-assisted linguistic research. The corpus must consist of authentic, naturally occurring data rather than computer-generated material, and its construction must be reflective, representing specific discourses and genres [Habert, B., 2000].

McEnery et al. Further specify the size, nature, and electronic usability of corpora: “A corpus is a large collection of authentic texts representative of a given variety of language and exploitable in electronic form” [McCarthy, M., McCarten, J., & Sandiford, H., 2006].

From these definitions, it follows that corpora must, on the one hand, include a sufficiently large number of varied texts to ensure representativeness, and on the other hand, be authentic - that is, drawn from real sources rather than artificially created for research purposes. In other words, the texts should not be authored by the linguists themselves in order to maintain objective representativeness.

Based on these definitions, Rastier (2002) outlines the specific features of texts and clarifies selection criteria as follows: a corpus is a structured grouping of complete, documented texts, possibly enriched through annotation, and assembled:

- Theoretically and reflectively, taking into account discourses and genres;
- Practically, with a view to a range of applications.

Thus, a corpus is constructed by selecting a subset of texts according to specific criteria for linguistic study [Rastier, F., 2002].

A corpus is inseparable from its linguistic use. In linguistics, it most often takes an electronic form - that is, a file or a set of digital files that can be processed using specific analytical tools. This digital dimension is essential, as it enables a range of analyses, such as extracting all occurrences of a given expression in context.

In speech therapy, however, corpora are typically analyzed manually to detect articulatory, phonetic, phonological, lexical, morphological, and syntactic errors. This manual examination requires going through all texts one by one, which can be quite time-consuming.

## **2) Corpora in the Study of Language**

Corpora are not a recent innovation in linguistics; they have always been an essential tool for the study of languages. Indeed, in order to analyze a language that one does not speak, it is necessary to rely on existing linguistic productions. Among the most famous ancient textual resources is the Rosetta Stone, which bore a decree of Ptolemy V dating from 196 BCE, written in Ancient Egyptian and Ancient Greek, with Egyptian appearing in two writing systems : hieroglyphic and demotic. This corpus proved decisive for Champollion, who used it in 1822 to decipher hieroglyphics. Without this exceptional document, the decoding of that writing system would have been impossible. Thus, for centuries, linguists have relied on textual resources to study not only ancient languages but also those with oral traditions [Parkinson, R. B., 1999].

However, these are not corpora in the modern sense of the term. These ancient linguistic resources were not generally compiled systematically to represent a particular text genre or language register. The systematic identification of textual concordances - extracts that allow words to be analyzed in their contexts of occurrence - is a relatively recent practice.

By the end of the 19th century, manual collections of occurrences already existed, particularly during the compilation of the *Oxford English Dictionary*. Nevertheless, it was not until the advent of computing in the 1960s that corpus linguistics began to

take the form we know today. Corpora at that time were still relatively small, but for English, representative resources already existed, such as the Brown Corpus for American English and the Lancaster-Oslo-Bergen (LOB) Corpus for British English.

In France, the Frantext Corpus represents a major literary corpus of the 19th and 20th centuries. It was created in digital form to serve as a documentary foundation for the *Trésor de la langue française*, a dictionary whose publication was completed in 1994. The development of corpus linguistics is therefore closely linked to that of computer technology.

With the rise of personal microcomputing in the 1980s, numerous large-scale corpora were created. The emergence of the Internet then accelerated this movement by facilitating access to corpora through specialized platforms and providing access to countless online textual resources. Whereas the first computerized corpora contained about one million words, today there exist massive resources such as the *Collins English Corpus*, which includes more than 4.5 billion words [<https://www.atilf.fr/ressources/frantext/>, consulted on 04/10/2025 at 20:30].

### **3) The Training of Students in Corpus Linguistics**

Training in corpus linguistics aims to make future speech therapists aware of the multitude of methodological choices that researchers and practitioners face when studying language productions. It highlights the impact of these choices on data construction and the resulting analyses.

Another goal of this training is to develop in students a keen sense of observation and a reflective attitude toward attested language productions. It seeks to encourage them to adopt a critical perspective on their own linguistic practices and their future professional interventions.

The program also provides practical initiation into methodological tools that may prove useful in their professional activity. Finally, it raises awareness of good practices related to the collection, processing, and, where applicable, dissemination of linguistic data.

Linguistics and speech therapy maintain a close relationship due to their common object of study: language. Since speech therapists regularly collect and analyze linguistic data from their patients, the relationship between speech therapy and corpus linguistics appears, as Witko (2010) puts it, to be a “natural” one.

The author emphasizes that corpus linguistics provides speech therapy with a dual contribution. On the one hand, it offers practitioners valuable theoretical and methodological tools for processing and analyzing the linguistic productions collected in their clinical practice. On the other hand, it constitutes a relevant research framework for studying language pathologies and improving their therapeutic management.

It is therefore unsurprising that the notion of the corpus has been integrated into the new training framework published in Algeria in 2025. This decision reflects the growing recognition of the central role played by authentic linguistic data in understanding, assessing, and managing language disorders. By integrating corpus linguistics into the Master’s program in *Language and Communication Pathologies*, the designers of the framework demonstrate their commitment to grounding speech

therapy training in an empirical and scientific approach to language, in line with recent international developments in the field.

The current curriculum includes, in this regard, a 1.5-hour weekly lecture devoted to corpus linguistics. This course is part of the so-called *transversal teaching units*, which highlights its bridging role between the linguistic, methodological, and clinical dimensions of the program. However, there is currently no practical tutorial (TD) specifically associated with this module. This absence raises questions, as the concrete manipulation of corpora and the analysis of authentic language productions are practical skills that cannot be fully developed through lectures alone.

Corpus linguistics, indeed, cannot be reduced to a set of theoretical concepts ; it requires active implementation involving the collection, cleaning, annotation, and analysis of real linguistic data. These essential steps ensure comprehensive training, allowing students to understand the methodological and ethical challenges of data handling, while also developing a critical perspective on professional and research practices in speech therapy.

From this standpoint, it seems essential to strengthen the role of this module in the curriculum. The addition of tutorial sessions would naturally complement the lecture. These sessions would give students the opportunity to experiment with corpus construction, use linguistic analysis tools, and apply good practices in data processing. They would thus promote a stronger link between theory and practice, while contributing to the professionalization and reflexivity of future speech therapists.

Here’s the **academic English translation** of your text:

**The program addresses the following key points:**

**A) Ethics of Data Collection**

The collection of oral or multimodal data constitutes an essential stage both in research and in clinical practice in speech-language pathology. However, it requires in-depth reflection, not only on scientific and clinical aspects, but also on legal and ethical considerations.

Before any oral or multimodal data collection, it is crucial to comply with a number of legal and ethical requirements. Two steps are indispensable:

1. **Informing the participant** about the framework within which the data collection takes place (scientific or clinical objectives, recording procedures, data processing and use, anonymization, storage, protection, dissemination, etc.)
2. **Obtaining their free and informed consent** in light of this information [Baude et al. 2006].

Introducing speech-language pathology students to the practice of oral or multimodal data collection from the very beginning of their training represents a privileged opportunity to raise awareness of these best practices. Indeed, they will encounter these issues repeatedly throughout their education - during case studies, in the context of their final dissertation, and later, in their daily clinical practice.

Even though data collection in speech-language pathology is governed by professional confidentiality, it must still adhere to precise rules regarding respect for consent and the protection of personal data.

Recording is, moreover, widely recognized as a “*natural*,” “*frequent*,” and even “*essential*” activity in speech-language pathology practice [Tran & Bécavin, 2001; Witko, 2010]. Such recordings make it possible to listen again to speech productions when writing assessment reports, to conduct detailed analyses of specific linguistic phenomena, or to review precise excerpts with the patient (and possibly their relatives). They also allow the preservation of spontaneous productions, complementing elicited ones.

According to Witko (2010), these recordings facilitate longitudinal observation of the patient’s language - from form to use - as well as analysis of interactional dynamics, both of which are crucial for adjusting therapeutic goals. They also serve as a **self-evaluation tool** for the clinician, enabling reflection on their own linguistic practices [de Weck, 2003; Rodi, 2011; da Silva, 2010], and contribute more broadly to scientific research on language disorders through the use of authentic data corpora.

However, these legal and ethical considerations are not without impact on clinical or scientific observation. As Baude et al. (2006) remind us, the very act of informing a subject about the objectives of a recording can alter the observed situation and compromise the spontaneity or “*naturalness*” of the collected data. This phenomenon is even more pronounced when the patient experiences significant linguistic insecurity, often linked to the very nature of their language difficulties [Tran & Bécavin, 2001].

Consequently, **establishing a climate of trust** between patient and clinician becomes essential to ensure the quality of the data collection. As Witko (2010) emphasizes, regular practice of recording facilitates its acceptance, particularly when it is presented as an ordinary professional tool - or even as a genuine participant in the interaction [Morgenstern, 2016].

## **B) The Use of Recording**

From a scientific standpoint, recording represents only the first technical step in building a corpus. It is embedded within a broader set of methodological and theoretical choices that must be clarified beforehand. These decisions directly influence the nature of the data collected, their interpretation, and their clinical or scientific significance. Training future speech-language pathologists to engage in this methodological reflection from the very beginning of their studies helps them develop the mindset of a practitioner-researcher, attentive to the quality, validity, and ethics of their observations.

It is during the data analysis stage that students can truly grasp the implications of the methodological choices made earlier. They concretely discover that these decisions - whether related to the type of data collected, the recording conditions, or the transcription method - directly affect the nature of observations and possible interpretations. This awareness enables them to understand why such parameters are typically defined *prior to data collection*, in accordance with scientific [Sinclair, 1996, cited in Habert et al. 1997] and/or clinical objectives [da Silva Genest & Masson, 2017]. It also leads them to realize that the various stages of corpus-based research are interdependent and must be conceived as part of a coherent whole [Adolphs & Knight, 2010].

As part of this training, students are asked to produce an audio or video recording (depending on parental authorization) of a kindergarten child describing a drawing they have just completed. This choice of task offers several pedagogical and methodological benefits. On the one hand, it exposes students to the real-life challenges of collecting oral data from young children and the difficulties some may experience speaking spontaneously in front of a microphone or camera [Blanche-Benveniste & Pallaud, 2001]. On the other hand, it provides an opportunity, during transcription, to address the issues raised by processing non-standard linguistic productions.

During the analysis phase, students are encouraged to question the **influence of their own presence - as well as that of the recording device** - on the observed productions, leading them to reflect on the well-known “*observer’s paradox*” [Labov, 1972]. This reflection represents a crucial step in preparing them for clinical practice, where the relationship between clinician and patient inevitably shapes linguistic productions.

No strict criteria are imposed for selecting the children to be recorded, in order to give students a degree of autonomy. Some choose to record children they perceive as “*good speakers*,” while others select those who seem to face difficulties. In both cases, transcription and analysis often yield unexpected findings: a child initially seen as having “*limited language skills*” may demonstrate fine linguistic competence, whereas one considered “*advanced*” may still display developing forms.

Students are encouraged to **interact with the child** during recording, in a dyadic situation similar to those they will encounter in speech therapy sessions. This setup allows them to adopt a **reflective stance** toward their own linguistic productions - whether structural, functional, or interactional in nature. The two-person exchange also facilitates later stages of transcription and analysis by reducing interactional complexity.

The choice of **drawing as an interactional support** reflects the intention to align data collection with other courses, particularly those focused on the development of graphic skills. This approach promotes **interdisciplinary dialogue** while stimulating the child’s verbal expression through a self-created medium. However, this situation presents certain **limitations for linguistic analysis**, especially when video recording cannot be performed: the numerous **deictic expressions** produced by the child often refer to visual elements absent from the audio recording. This observation leads students to revisit earlier discussions about the **advantages and limitations of various data collection methods** [Behrens, 2008].

Despite these constraints, the experience serves as a **valuable introduction to the collection and processing of spontaneous productions**, which Rondal (2003) advocates as an essential component of speech-language assessment. During analysis, instructors emphasize the **partial nature** of the data collected: it represents only a fragment of the child’s linguistic abilities. Students are thus made aware of the need for **interpretive caution** and the importance of avoiding unwarranted generalizations in their analyses.

No specific verbal instructions are provided to guide the child, allowing students to question the **influence of the given instruction** on the type of discourse produced [Blanche-Benveniste & Pallaud, 2001]. The **minimum duration** of the recording is set at **three minutes**. Although time is not a particularly relevant measure in corpus linguistics [Baude et al. 2006, p.28], this time constraint provides students with a reassuring framework. However, it can become a source of difficulty when the child's speech rate is slow - offering another pedagogical opportunity to discuss the **size and representativeness** of a corpus, depending on the linguistic phenomena under study.

Thus, this activity not only introduces students to the **fundamental principles of corpus collection and analysis**, but also initiates them into a **reflective approach** toward their future professional practices.

The analysis of the collected corpora also provides an opportunity to remind students that a corpus does not need to be large to be meaningful. As Tomasello and Stahl (2004) emphasize, the **density of a corpus** is not determined solely by the amount of speech produced, but by the **richness of the information it contains**. Even when a child speaks little, numerous observations can be drawn from their productions - regarding the structure of utterances, the communicative strategies employed, and the interactions between verbal and nonverbal language. The studies of Debras (2018) and Canut & Vertalier (2008) further demonstrate that **in-depth qualitative analyses** can be conducted on limited individual productions, provided that the observation is fine-grained and contextualized.

This approach mirrors the **realities students will face in their future clinical practice**. In speech-language pathology, it is common to encounter patients - children or adults - whose verbal productions are restricted, fragmentary, or even absent. Such situations do not represent an obstacle to linguistic assessment; on the contrary, they call for heightened attention to **nonverbal manifestations of language**: gestures, gaze, facial expressions, body posture, and interactional cues that reveal comprehension or communicative intent. By working on **low-density corpora**, students learn to identify the **informative value of each fragment** and to interpret communicative behaviors holistically.

For this activity, students are asked to make their own recordings of a child's productions using their **mobile phones**. This choice is deliberate : it aims to place students in **data collection conditions similar to those encountered in the field**, while reducing technical demands. The mobile phone is a familiar everyday tool that allows them to focus on the **quality of the interaction** rather than on handling a complex recording device. Moreover, children today are generally **comfortable with the presence of such devices**, which helps make the situation more natural. As Habert (2004) notes, using a **simple and familiar device** reduces the anxiety sometimes associated with technology and promotes a **more authentic interaction**. The goal is not merely to simplify the task but to show students that an **accessible tool** can be effectively used in their future practice, provided that **ethical and methodological principles** of data collection are respected.

Once the recordings have been made, an essential step of the work consists in encouraging students to **critically reflect on their own methodological choices**. This

reflective pause allows them to become aware of the **active role of the observer** in the construction of a corpus. They are thus led to question several aspects: the moment chosen to start the recording (did they record the instruction given to the child or only the response?), the elements included in the frame in the case of a video recording (did they film only the child, or also the drawing being made? did they choose an angle that included the adult?), or the duration and continuity of the recording (did they interrupt or continue the recording depending on the dynamics of the interaction?). These decisions - sometimes made unconsciously - nonetheless influence the **nature and interpretation of the data**.

The collective analysis of these practices during the course highlights the **importance of context in language study**. Language productions cannot be understood independently of the **interactional framework** and the **nonverbal behaviors** that accompany them. Visual or prosodic cues are often necessary to correctly interpret certain forms or silences. This awareness is fundamental, as it leads students to conceive of language acts as **multimodal and situated phenomena**, rather than as simple sequences of words.

Finally, in preparing their analyses, students are invited to examine the **strengths and limitations** of their approach and to propose an **alternative methodology**. In doing so, they learn to identify potential biases, justify their choices, and suggest improvements. This **critical process** constitutes an essential aspect of developing the **researcher-practitioner mindset**: it prepares them to design, evaluate, and adjust their methods in future clinical or research situations. Through the practice of observing, analyzing, and questioning their own approaches, students develop a **reflective competence** that will be indispensable throughout their professional careers as speech-language pathologists.

The **transcription of oral productions** represents a decisive stage in training for linguistic and clinical analysis. One of the first observations made with students concerns their tendency to attribute to the child certain characteristics of ordinary spoken French: elisions, dislocations, omission of the negative particle *ne*, hesitations, repetitions, reformulations, or filled pauses. These phenomena - typical of spoken French - are sometimes mistakenly interpreted as indicators of linguistic immaturity or speech difficulties. This interpretive bias reveals a crucial issue: the **reference model** the student adopts when evaluating another person's speech. Do they rely on a **written, standardized, and stable** model of language, or on a **spontaneous, contextualized conception** of spoken language? This question, central to the teaching of **applied linguistics for speech therapy**, invites an **epistemological reflection** on the very nature of language and the diversity of its forms.

To help students move beyond these **normative representations**, they are asked to **transcribe their own speech**, taken from real or simulated interactions. This exercise - often disconcerting - gives them firsthand experience of the **complexity of so-called “standard” spoken language**. They thus discover that disfluencies, ellipses, and repetitions are not anomalies, but **natural manifestations of language in action**, constrained by interactional and cognitive factors. By recognizing these phenomena in their own speech, they develop a **more sensitive and less judgmental ear** toward

the productions of others, whether children, adults, or patients with language disorders.

This **introspective process** also leads them to understand that “speech accidents” – [Tran & Bécavin, 2001] are **intrinsic features of healthy spoken language**, and that it is essential to distinguish between **typical oral variation** and **pathological signs**. As Loufrani (2000) points out, the speech-language pathologist must interpret a subject’s productions not according to an **abstract linguistic ideal**, but in relation to the **communication context** and the **conditions of the given task**. Thus, a given utterance might be considered “atypical” in a school setting, yet perfectly expected in a spontaneous or emotional exchange.

Analyzing their own transcriptions leads students to reflect on the **frequency and regularity** of phenomena they had previously perceived as marginal. The systematic nature of transcription - which requires noting every word, pause, and repetition - reveals the **recurrence of structures** such as dislocations and incomplete utterances. This awareness plays a **formative role**, helping to deconstruct the idea of spoken language as a “deficient” version of writing and to adopt a **more realistic view of linguistic competence**. Such a perspective is crucial for clinical practice: supporting a patient means helping them to develop a **functional and socially appropriate use of language**, not to imitate an artificial norm. The speech-language pathologist thus becomes a **mediator between situated linguistic practices and the communicative demands** of the patient’s social environment [Tran & Bécavin, 2001].

Furthermore, transcription creates a space for reflection on how students conceptualize syntax and its acquisition. Depending on the conventions adopted, transcription may include specific annotations to indicate fixed constructions, routine formulas, or multiword expressions produced by the child. This annotation work makes visible the gradual and dynamic nature of syntax, as envisaged by constructivist models. Behrens (2008), for example, shows that highlighting these fixed segments sheds light on how the child gradually constructs syntactic schemas from preformed units. This perspective can be linked to Goldberg’s (1995) work on construction grammar, which posits that linguistic competence is built upon the accumulation and generalization of expressions grounded in actual language use.

These theoretical approaches have direct implications for speech-language pathology practice. Understanding that a child’s productions often stem from partial or routinized combinations rather than from poorly applied abstract rules enables practitioners to adopt a more nuanced approach to assessment and intervention. In the field of language disorders, this view is supported by Leroy (2013), who suggests that certain syntactic “errors” should not be seen as deficits, but as indicators of the constructive functioning of language in development or reorganization.

Finally, this reflective transcription experience can be extended through a discussion on the role of metadata in corpus construction. In attempting to understand why a given phenomenon appears or recurs, students are encouraged to formulate explanatory hypotheses that require contextual information - such as the speaker’s age, communicative situation, type of interlocutor, duration of the exchange, topic, and degree of familiarity. This reflection on external variables paves the way for a

truly scientific approach, in which transcription is no longer viewed as a mere technical step, but as a tool for observation and analysis embedded within a rigorous methodological framework.

### **C) Transcription**

Once the recording has been completed, transcription becomes a central step in both the construction of a corpus and the training of speech-language pathology students. Because the process of transcription can be long, meticulous, and intellectually stimulating, students are invited to work in pairs. After each has listened to or viewed their respective recordings, they select the one they consider most relevant and transcribe it individually. When comparing their transcriptions, they often realize that they have not perceived the same elements [Cappeau, 2008]. This divergence in listening leads them to pay renewed attention to the data and to question the subjectivity of auditory perception, their own representations of language and speech, as well as their expectations regarding the recorded participant. They also notice that they have not always represented the same observable features in the same way, prompting reflection on the implicit choices involved in any act of transcription [Ochs, 1979].

Because it requires multiple listenings or viewings, transcription often reveals details that students did not notice during the first observation - such as a gesture accompanying an utterance, a facial expression, or a nonverbal reaction from the interlocutor. These discoveries encourage them to reflect on their conception of language and the role they assign to nonverbal aspects in interaction.

At the same time, students become aware of their initial expectations toward the recorded subject. For instance, some who believed they had selected a child “who doesn’t speak very well” realize during transcription that, despite certain phonological features that affect intelligibility, the child produces syntactically complex utterances. This realization invites them to reflect on the nature of their listening, which is often oriented toward detecting pathological indicators [see Tran & Bécavin, 2001: 31]. Through transcription and corpus analysis, this “biased” listening can be confronted with a set of measurable and contextually grounded observations within the target subject’s linguistic system, thereby enhancing the credibility and rigor of qualitative analysis [Witko, 2010: 18].

In this context, students often attribute to children’s speech certain features that in fact belong to ordinary spoken French: elisions, dislocations, omission of the negative *ne*, hesitations, repetitions, or lexical searches. This observation serves as a valuable starting point for encouraging them to question the reference model they adopt when analyzing children’s productions: do they rely on child-directed speech, or on a more normative written model? Transcribing their own productions allows them to become aware - “from the inside” - of the specific features of standard, living, and natural spoken French. They also discover that the so-called “speech accidents” [Tran & Bécavin, 2001] are an integral part of non-pathological spoken language, and that, in analyzing patients’ productions - whether children or adults - it is essential to distinguish between phenomena that stem from a disorder and those that simply reflect the oral modality of a communicative task [see Loufrani, 2000].

The analysis that follows transcription leads students to identify possible explanatory factors for the observed phenomena. By reflecting on external parameters, they come to understand the importance of collecting metadata during data gathering - such as the communicative situation, interactional context, or degree of familiarity between interlocutors - in order to formulate sound hypotheses.

The systematic nature of transcription also makes it possible to assess the true frequency of certain phenomena, such as dislocations, which are often underestimated. Questioning their representations of spoken French thus becomes a key stage in their training. In their future professional practice, speech-language pathologists will guide patients toward a *realistic* use of language - one adapted to their social and communicative environment - rather than toward a strictly normative model [Tran & Bécavin, 2001].

Transcription also leads students to rethink their understanding of syntax and its acquisition. Depending on the conventions adopted, it may include the annotation of potentially fixed constructions produced by the child [Behrens, 2008]. This approach makes it possible to introduce constructivist hypotheses applied to adult syntax [Goldberg, 1995] and to language pathology [Leroy, 2013], showing that linguistic development is based on the progressive combination of preformed units derived from use and experience.

Just as many methodological choices are made prior to data collection, transcription itself involves a series of preliminary decisions - concerning tools, conventions, and the level of detail retained - all closely guided by the theoretical framework adopted and by the type of analyses envisaged (qualitative, quantitative, or mixed). These technical choices are never neutral: they reflect a particular conception of language and directly influence the observable features that can be extracted from the corpus.

Transcription also confronts students with the delicate question of interpreting the target speaker's productions, particularly when the transcriber and the speaker do not share the same language variety, or when the speaker is a child still acquiring their native language [Morgenstern & Parisse, 2007]. This situation is all the more frequent in cases of language disorders, where the production may differ significantly from expected norms [Tran & Bécavin, 2001].

When transcribing using standard orthography, students are often led to *interpret* the subject's productions - for example, deciding whether a sound sequence corresponds to a poorly articulated word, a child-like form, or an unintelligible segment. Conversely, they may choose to suspend interpretation when uncertainty remains. These conscious or unconscious choices have a direct impact on the subsequent linguistic and developmental analyses.

To support this interpretative dimension of the work, it is essential to create spaces for reflection and discussion - in class, in workshops, or on online forums - where students can make explicit the difficulties they encounter during transcription. These collective exchanges encourage the use of linguistic concepts and terminology acquired in previous teaching units, while fostering the development of an analytical stance.

Articulating the difficulties they face leads students to produce precise, non-prescriptive descriptions that demonstrate a genuine awareness of the specific features of spoken language and of the child’s speech. This reflective work contributes to building a critical and detached perspective on the various choices made, thus representing a key formative step in learning the scientific approach within speech-language pathology.

#### **4) The Use of the Corpus in Speech-Language Therapy Practice**

Speech-language therapy is currently experiencing a particularly dynamic expansion in Algeria. Although this discipline is relatively new and still little known to the general public, it is gradually becoming more structured - supported by the creation of new university programs, the diversification of teaching modules, and the establishment of better-supervised clinical internships. These developments reflect an institutional and scientific will to meet a real need: that of specialized care for language and communication disorders in a country marked by complex linguistic and cultural realities.

Indeed, Algeria is characterized by a plural linguistic landscape where several languages coexist: dialectal Arabic (*daridja*), used in everyday life; the Berber languages, including Kabyle, which are deeply rooted in certain regions; and French. This situation of multilingualism and diglossia directly influences children’s language development, family communication dynamics, and, consequently, the practice of speech-language therapy.

For the speech-language therapist, this linguistic diversity represents both a richness and a challenge. It requires constant adaptation to the language of the patient and their environment, in order to ensure relevant assessment and effective intervention. A language disorder can only be understood by taking into account the language or languages in which it is expressed. Consequently, a monolingual approach or one modeled on imported frameworks risks producing biased assessments.

In this context, the use of linguistic corpora emerges as an essential tool. The corpus - understood as a structured collection of language productions gathered in natural contexts - makes it possible to observe and analyze language as it actually manifests in the patient, rather than relying solely on standardized tests. It offers a more nuanced and contextualized view of language abilities, taking into account interactions, social codes, and linguistic variations specific to each speaker.

The advantages of corpus use in speech-language therapy practice are numerous :

1. **Authentic observation of language:** the corpus allows for the capture of spontaneous productions, hesitations, reformulations, and gestures accompanying speech, thereby providing access to the real dynamics of communication [see Habert, 2004].
2. **Better differentiation between disorder and linguistic variation:** In a multilingual context, corpus analysis helps distinguish what belongs to the normal functioning of a language or to language contact phenomena, and what corresponds to an actual language disorder [see Tran & Bécavin, 2001].

3. **A diagnostic and developmental monitoring tool:** By comparing productions over time, the corpus makes it possible to evaluate the patient's progress, adjust therapeutic goals, and document the effects of interventions.
4. **A support for reflection among students and practitioners:** The creation and analysis of corpora promote the development of a scientific, critical, and reflective attitude. Speech-language therapists learn to observe, transcribe, code, and interpret linguistic data using rigorous methodologies, while becoming aware of their own representations of language [see Ochs, 1979].
5. **A lever for local research:** Oral corpora built in the Algerian context can feed into national databases, useful both for linguistic research and for the development of assessment tools adapted to local languages [Darija, Kabyle, Chaoui, etc.].

Despite these advantages, the use of corpora remains only weakly integrated into speech-language therapy practice in Algeria. In the field, it is observed that few therapists use them systematically - often due to a lack of specific training, adapted resources, or time dedicated to such an approach. Most assessments still rely on standardized tests designed for other linguistic contexts, which can lead to misinterpretations and limit the relevance of diagnoses.

To address this, it would be desirable to give greater importance to corpora in the university training of future speech-language therapists. This would involve not only theoretical instruction on corpus building, transcription, and analysis, but also practical workshops where students would learn to collect and analyze language data from patients or children in natural situations. Such training would foster the development of fine observational skills, essential for distinguishing linguistic markers of disorder from simple variations inherent to the patient's language.

### **Conclusion**

In conclusion, the introduction of corpus linguistics into the training of speech-language therapists in Algeria represents a decisive step toward a more professional practice, firmly grounded in both research and the linguistic reality of the country. However, for this progress to become truly effective, it is essential to strengthen its pedagogical and methodological foundations. Integrating tutorials, transcription workshops, and data analysis sessions would help transform theoretical knowledge into genuine practical competence - an indispensable skill for a nuanced understanding of language production.

The corpus is not merely an analytical tool; it also serves as a means to cultivate a reflective, critical, and context-sensitive stance toward language. By promoting an empirical and scientific approach, it enables future speech therapists to adapt their interventions to Algeria's linguistic diversity, refine their diagnostic assessments, and contribute to the development of local scientific knowledge on language functioning in multilingual contexts.

Thus, fostering a true corpus-based culture in speech-language pathology is to work toward more demanding training, more rigorous clinical practice, and research that is deeply rooted in the Algerian field. It also paves the way for the construction of an authentically Algerian speech-language therapy - capable of engaging with

international standards while remaining attentive to the cultural and linguistic specificities of its environment.

Promoting a corpus-based culture in speech-language therapy training and practice in Algeria means paving the way for a more contextualized, scientifically rigorous, and linguistically inclusive approach. This would enhance diagnostic quality and optimize patient care, while also contributing to the development of local scientific knowledge about the specificities of language in a multilingual context.

## 7. REFERENCES

1. Aldolphs, S., & Knight, D. (2010). Building a spoken corpus: What are the basics? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 38–51). Routledge. (in French)
2. Baude, O., Blanche-Benveniste, C., Calas, M.-F., Cappeau, P., Cordereix, P., Goury, L., ... Mondada, L. (2006). *Spoken corpora: Guide to good practices 2006*. CNRS Editions; Presses Universitaires Orleans. (in French)
3. Behrens, H. (2008). *Corpora in language acquisition research: History, methods, perspectives*. John Benjamins Publishing. (in English)
4. Blanche-Benveniste, C. (2010). *Approaches to spoken French language*. Editions Ophrys. (in French)
5. Blanche-Benveniste, C., & Pallaud, B. (2001). Collection of children’s utterances: Recordings and transcription. *Recherches sur le francais parle*, 1(16), 11–37. (in French)
6. Cappeau, P. (2008). Establishing spoken corpora: What transcriptions teach us. *Verbum*, 30(4), 343–353. (in French)
7. Canut, E., & Vertalier, M. (2008). Representative data... of what in language acquisition? Construction of observational data and analytical objectives. *Verbum*, 30(4), 299–312. (in French)
8. Da Silva Genest, C., & Masson, C. (2017). The contribution of corpus linguistics to the study of clinical situations: The use of ecological resources. *Studii de Lingvistica*, 7, 89–112. (in French)
9. Da Silva, C. (2011). Dialogic orientations in speech therapist–dysphasic child interaction. In *International conference “Dialogism: Language, discourse”*, September 2010, Montpellier. (in French)
10. Debras, C. (2018). Small and large corpora in linguistic analysis of gestures. *Corpus*, 18. Retrieved from <http://journals.openedition.org/corpus/3287> (in French)
11. De Weck, G. (2003). Language practices, interaction contexts and discourse genres in speech therapy. *Revue Tranel (Travaux neuchatelois de linguistique)*, 38, 25–48. (in French)
12. Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press. (in English)
13. Habert, B. (2004). Equipping linguistics: From borrowing techniques to encounters of knowledge. *Revue française de linguistique appliquée*, 9(1), 5–24. (in French)

14. Habert, B. (2000). Representative corpora: Of what, for what, how. *Cahiers de l'Universite de Perpignan*, 31, 11–58. (in French)
15. Habert, B., Nazarenko, A., & Salem, A. (1997). *Corpus linguistics*. A. Colin. (in French)
16. Leroy, S. (2013). Generalization disorders in construction grammars in children with specific language impairment (Doctoral dissertation). Ecole doctorale Connaissance, langage et modelisation, Nanterre. (in French)
17. Loufrani, C. (2000). Exploitation of data from pathological corpora. In M. Bilger (Ed.), *Corpus: Methodology and linguistic applications* (Vol. 3, pp.168–179). Honore Champion. (in French)
18. McCarthy, M., McCarten, J., & Sandiford, H. (2006). *Touchstone* (Textbook series, p. 5). Cambridge University Press. (in English)
19. Morgenstern, A. (2016). Language practices and patient behavior in the family environment: Contribution of multimodal ethnographic methods for medical research. *Ethics, Medicine and Public Health*, 2(4), 641–649. (in French)
20. Morgenstern, A., & Parisse, C. (2012). The Paris corpus. *Journal of French Language Studies*, 22, 7–12. (in English)
21. Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp.43–72). Academic Press. (in English)
22. Parkinson, R. B. (1999). *Cracking codes: The Rosetta Stone and decipherment* (pp.12–27). British Museum Press. (in English)
23. Rastier, F. (2002). Epistemological issues of corpus linguistics. In *Corpus linguistics* (pp. 31–45). (in French)
24. Rodi, M. (2011). Speech therapist–child interactions: How potentially acquisitional exchanges are constructed in a symbolic play situation. *Glossa*, 110, 49–68. (in French)
25. Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language*, 31(1), 101–121. (in English)
26. Tran, T. M., & Becavin, A.-S. (2001). Evaluating oral expression: Reflection on speech therapy practice. *Glossa*, 78, 22–33. (in French)
27. Witko, A. (2010). Corpus and language pathology: What approach in speech therapy? *Cahiers de Praxematique*, 54–55, 119–144. (in French)