CNRS - Toulouse INP - UT3 - UT Capitole - UT2

Institut de Recherche en Informatique de Toulouse

Traitement Automatique de la Parole & Intelligence Artificielle

Petite cartographie des enjeux techniques, scientifiques et éthiques

Table ronde – SAES - 6/06/2025

Isabelle Ferrané – isabelle.ferrane@irit.fr

TAP & IA – d'abord une question de point de vue

RECHERCHE

Vue « Observateur »

Analyse du contenu d'enregistrements audio
(off line)
ou de flux audio
(on line)
dans un but précis

USAGES

Vue « Interlocuteur »

Parler, converser oralement avec à une machine

… d'égal à égal ?

Isabelle Ferrané – isabelle.ferrane@irit.fr

TAP & IA – enjeux côté Recherche

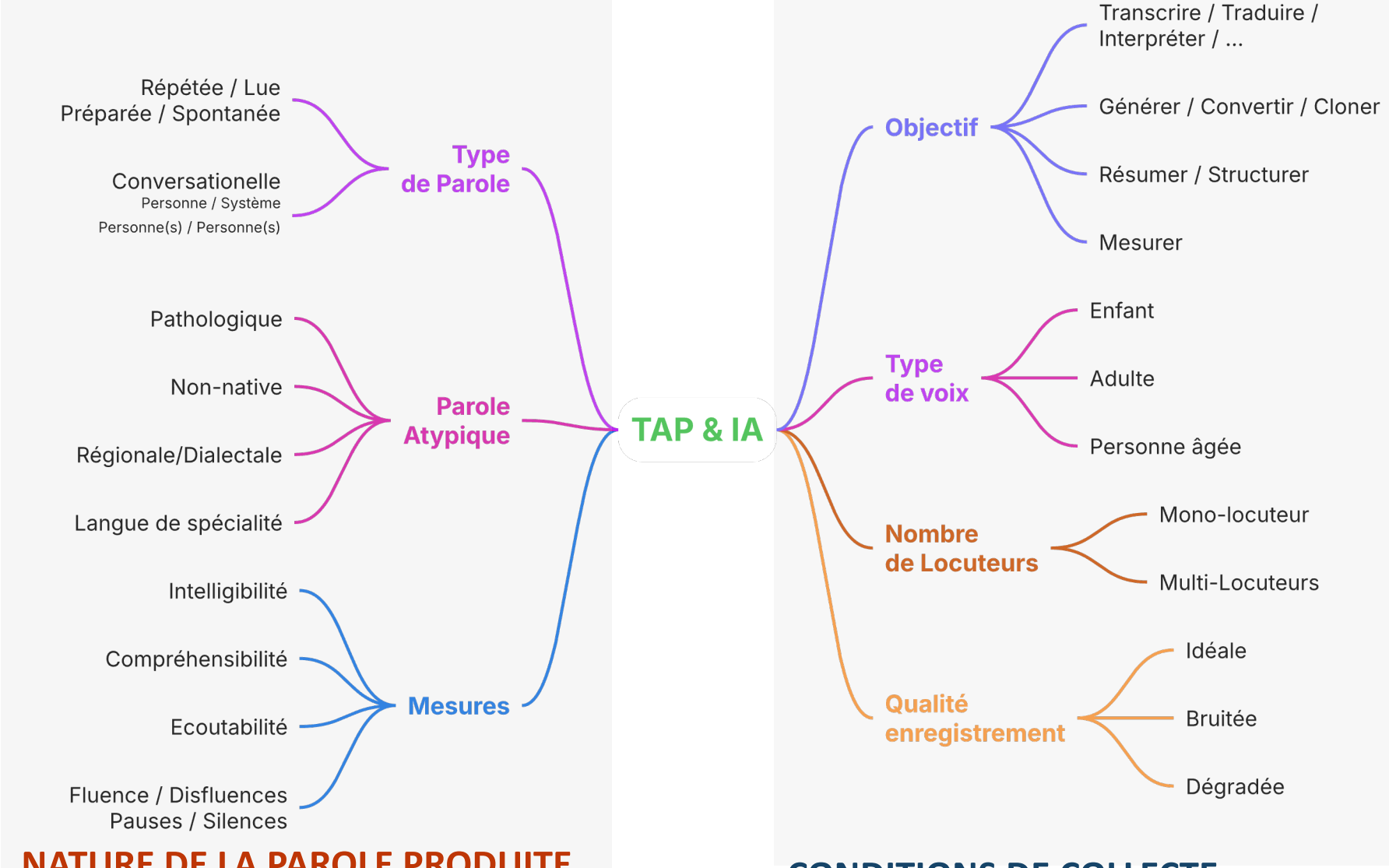
Traiter de la parole = Pour quoi faire ? Pour quels objectifs ?

Objectif	Entrées → Sortie
Transcrire (STT) / Traduire (STS)	Parole → Texte / Parole (Langue Cible)
Interpréter (NLU / SLU)	Parole → représentation du sens
Générer (TTS) Convertir Cloner	Texte → Parole Voix Source → Voix Cible même texte Texte → Voix Cible
Résumer / Structurer	Discussion → Texte + Structure
Mesurer	Parole / Voix → métriques et valeurs associées (globales/locales)
…	…

Isabelle Ferrané – isabelle.ferrane@irit.fr

TAP & IA – enjeux côté Recherche

Cartographie



Isabelle Ferrané – isabelle.ferrane@irit.fr

TAP & IA – rôle de l'IA

Présente depuis le départ

1950 : Alan Turing : *une machine peut-elle penser ?*

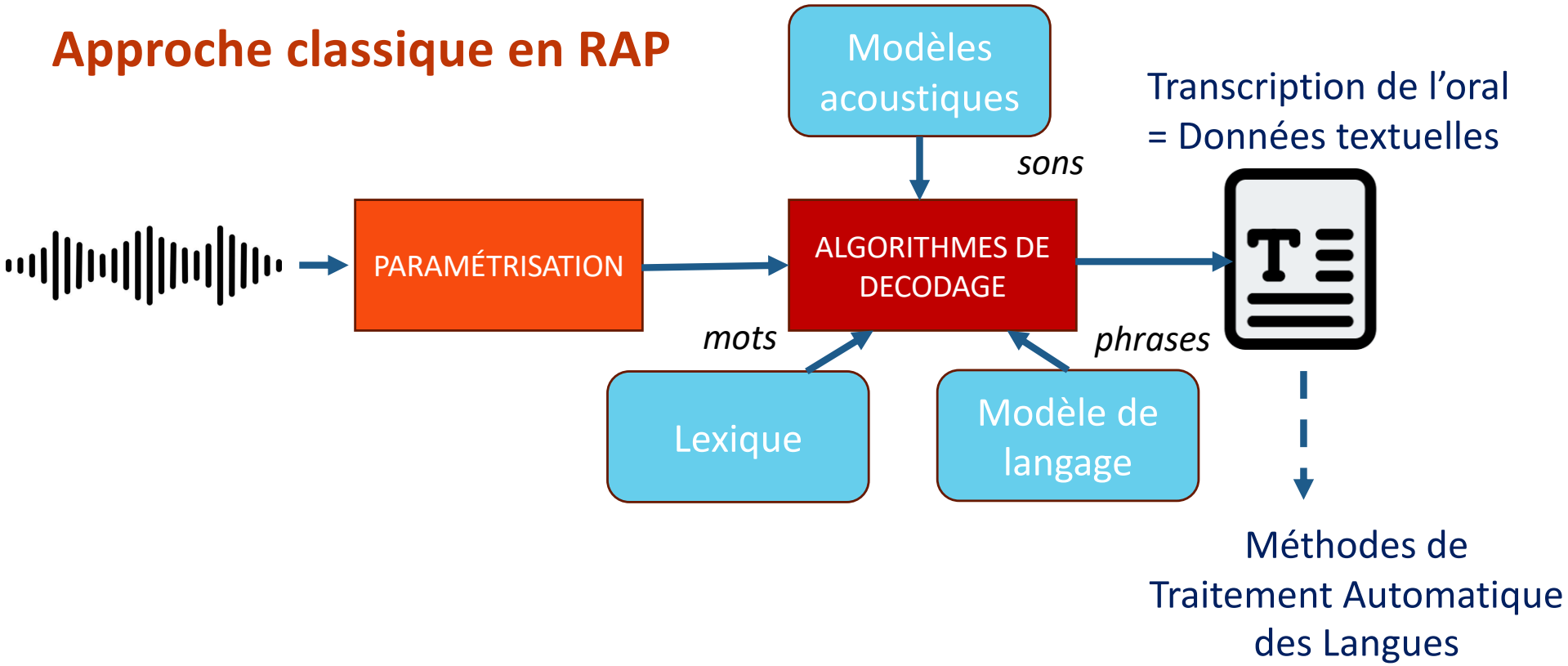
➔ Test = évaluer la capacité d'une machine à imiter la conversation humaine.

Intelligence Artificielle : quelles approches ?

Approches	Méthodes
Symbolique	Orientée représentation logique et raisonnement
Reconnaissance des formes	Basée sur l'apprentissage automatique et les statistiques Apprentissage supervisé / non supervisé
Connexionniste	Inspirée du fonctionnement du cerveau humain Réseaux de neurones, Perceptron multicouche, … …. Réapparue en force 2012 avec **l'apprentissage profond** : vision par ordinateur, génération d'images, **reconnaissance de la parole, synthèse vocale, traduction automatique, …**

TAP & IA – Impact sur le TAP

Approche classique en RAP



Paramètres à extraire directement du signal (temps, fréquence, …)
Modélisation des connaissances de la langue : phonèmes, lexique, syntaxe
Prédire un mot connaissant les mots précédents

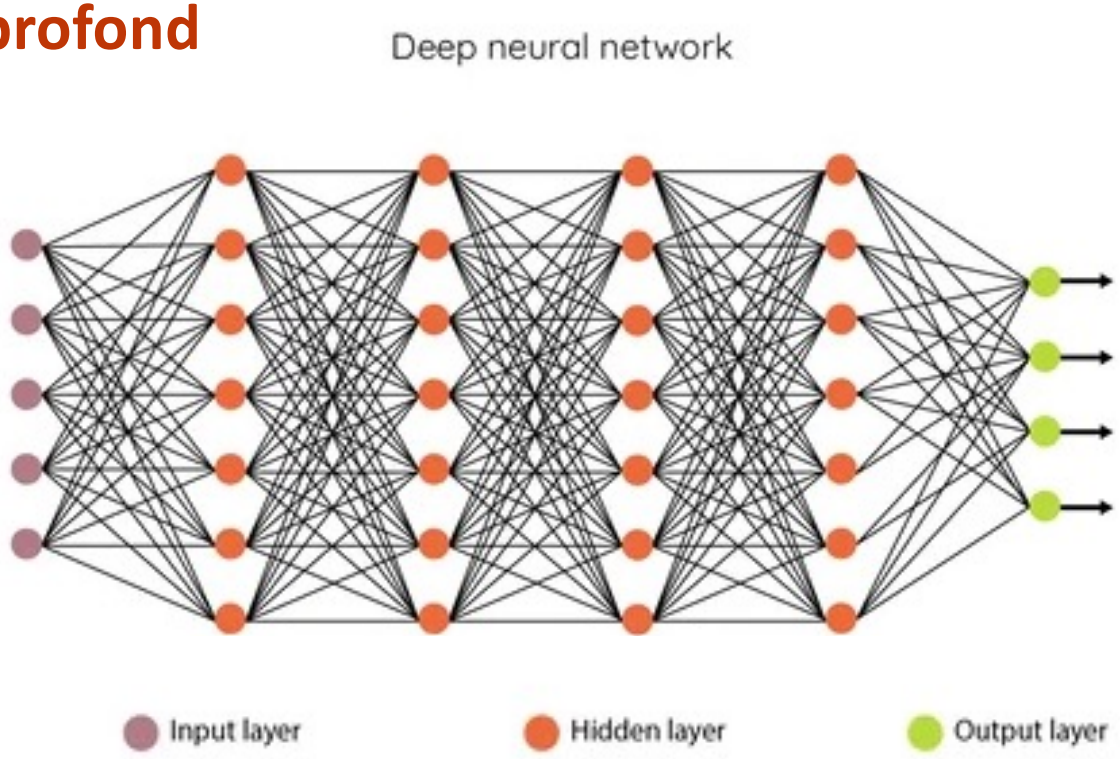
→ **Contexte limité mais les résultats des différentes étapes sont accessibles**

TAP & IA – Impact sur le TAP

Apprentissage profond

Représentation vectorielle (numérique) en entrée

x_1
x_2
...
x_d



Résultats
Prédiction
Classification
Compression
Encodage
Décodage
...

REPRÉSENTATIONS
vecteurs, contexte, …
Embeddings, …

ARCHITECTURES
couches, neurones,
Transformer, …

MÉTHODES APP.
supervisée, non sup.
auto-supervisée

MODÈLES
appris,
adaptés

TÂCHES
transcrire
…

TAP & IA – Impact sur le TAP

Modèles pré-entrainés pour l'encodage des données

- Volume des données d'apprentissage : >> 100 Go texte ou 90 000 h audio
- Contexte très large : plusieurs centaines de tokens voire plus
- Coût d'entrainements : en milliers, millions d'heures
- Nombre total de paramètres (poids) : de l'ordre du million voire du milliard
- Tâche d'entraînement auto-supervisée : masquage de tokens
- Adaptation des modèles à d'autres tâches.

Familles de modèles à disposition (mono ou multilingue):

BERT (mots), SentenceBert (phrases), WavLM (audio), …

→ représentations contextualisées des données traitées

Inconvénients :

Opacité des modèles → manque d'explicabilité / d'interprétabilité

Tâche de transcription/génération → hallucinations

Généricité des modèles → non adaptés voix / spécialités

Utilisation en local → **la voix est une donnée personnelle (CNIL)**

https://www.dilepix.com/en/blog/difference-artificial-intelligence-deep-learning-agriculture

TAP & IA – côté USAGES

Interaction Personne/Système

Chatbots, VoiceBots, Assistants vocaux

- Interagir avec une personne
- Interpréter ce qui dit (NLU, SLU) ➔ Commande vocale vs conversation
- Prédire la prochaine action du système
  - aller au-delà du question-réponse = Gérer le dialogue (DM)
  - garder trace de ce qui a été dit/compris/fait = Gérer l' historique (Tracker)
- Générer une réponse (NLG, TTS) / se connecter à une API (mail, météo, …)

IA génératives

- basées sur les LLM (GPT, Gemini, …)
- tout parait magique … mais garder à l'esprit que :
  - vos données sont réutilisées,
  - le résultat produit est questionnable (hallucinations, sources, biais ?)

➔ **il n'est pas interdit de réfléchir et …**

d'avoir recours …. à l' *« intelligence naturelle »*