

# Traduction automatique et assistée

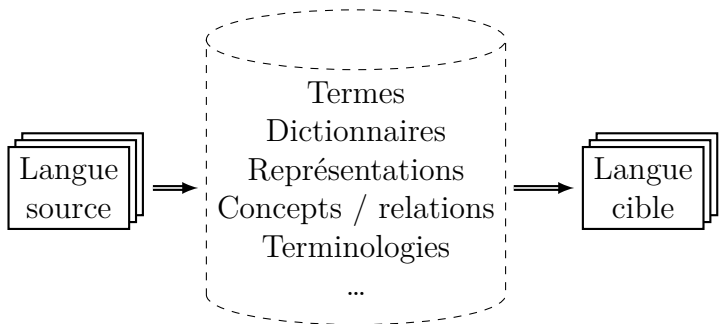
Damien Nouvel



# Plan

1. Automatiser la traduction
2. Traduction automatique
3. Traduction assistée

# Modèle pour la traduction



# Historique

- ▶ Quelques dates clé
  - ≈ 1799 : Pierre de Rosette
  - ≈ 1890 : Esperanto
  - ≈ 1930 : Machine de Trojanski
  - ≈ 1946 : Traduction électronique (UK)
  - ≈ 1955 : NUDE
  - ≈ 1960 : Triangle de Vauquois
  - ≈ 1970 : Systran
  - ≈ 1980 – 1990 : EUROTRA
  - ≈ 1985 : Reverso (Softissimo)
  - ≈ 2000 : Corpus Europarl
  - ≈ 2004 : Google translation service
  - ≈ 2010 : Linguee

# Complexités de la traduction

- ▶ **Diachronie** : langues / dialectes / variantes
  - ▶ Deux dimensions de difficultés
    - Trouver des correspondances entre termes (**paradigmes**)
      - Homonymes / synonymes
      - Tournures et emplois
      - Entités nommées (translittération)
    - Constructions (**syntagmes**)
      - Syntagmes / expressions polylexicales
      - Constructions syntaxiques
      - Prise en compte de la pragmatique (discours) ?
  - ▶ Deux méthodes informatisées pour la traduction
    - Traduction **automatique** et post-édition
    - Traduction **assistée** (mémoires de traduction)
- ⇒ Degré d'**intervention humaine**

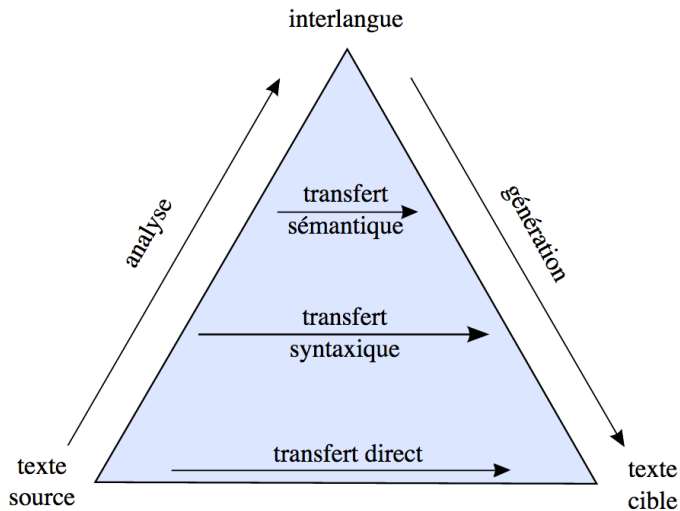
# Les approches

- ▶ Selon le **domaine**
  - Vocabulaire contrôlé (météo, transports, etc.)
  - Langue générale
  - Langue de spécialité (terminologie)
- ▶ Selon la **méthode** de traduction
  - Paires de langues ( $\approx 150 * 149 = 22K$ )
  - Langue pivot (carré :  $\approx 150 * 2 = 300$ )
- ▶ Des **corpus** comme **ressources**
  - **Parallèles** : alignement des phrases (bitext)
  - **Comparables** : alignement des sections / documents

# Plan

1. Automatiser la traduction
2. Traduction automatique
3. Traduction assistée

# Triangle de Vauquois





# Traduire segment à segment

- ▶ Impossible de tenir compte de tout le texte ...
- ⇒ Il faut **segmenter** / structurer
  - Énoncés
  - Propositions
  - Syntagmes
  - Mots
- ⇒ Meilleure hypothèse de traduction par recherche de segments
- ⇒ Difficultés avec les **dépendances** (accords, anaphores, etc.)

# Exploitation de ressources

- ▶ Chaque ressource a ses avantages et inconvénients
  - Lexiques bilingues
  - ⇒ Traduction **mot-à-mot** et ambiguë
  - Corpus parallèles
  - ⇒ Précis mais peu **couvants**
  - Corpus comparables
  - ⇒ Plus couvrants
  - Analyses linguistiques (morpho-syntaxe, syntaxe, sémantique)
  - ⇒ Vérification des constructions (accords, etc.)
- ⇒ Il faut **réordonner** les mots (réagencer les phrases)
- ▶ Moteurs de traduction
  - Par **règles** spécifiées par des humains
  - **Statistiques** à partir des données (corpus)
  - **Hybrides** (combinaison d'approches)

## Queqlues moteurs en ligne

- ▶ Google Traduction : <https://translate.google.fr>
- ▶ DeepL (Linguee) : <https://www.deepl.com/translator>
- ▶ Reverso (Soft.) : [www.reverso.net/text\\_translation.aspx](http://www.reverso.net/text_translation.aspx)
- ▶ Systran : <https://demo-pnmt.systran.net>

# Plan

1. Automatiser la traduction
2. Traduction automatique
3. Traduction assistée

# Approches

- ▶ Traduction automatique
  - **Post-édition** (correction)
    - ⇒ Quelle efficacité, quels biais de traduction ?
  
- ▶ **Mémoires de traduction**
  - Assistants de rédaction pour la traduction
  - Plus ou moins **personnalisés** / **mutualisés**
  - Approche par **segments** (séquences de mots)
  - Contrôle de la **cohérence** des traductions
  - ⇒ Logiciels professionnels (SDL Trados)