



École Pratique
des Hautes Études

Lutin Userlab
Cité des sciences et de l'industrie



Holken Consultants
& Partners



SYSTRAN
beyond language

PROJET « SIMPLES »
SIMplification des Langues EcriteS

NATALIA SEGAL | 12/11/2019

PROJET « SIMPLES » : ENJEUX ET OBJECTIFS

sociétal : simplification et résumé de contenus numériques complexes

- pour les personnes déficientes intellectuelles et leurs accompagnateurs
- pour les enfants, les apprenants de la langue, les personnes âgées... et tout le monde ! ☺

scientifique et technique : logiciel combinant les règles et l'apprentissage automatique

- un véritable défi à la communauté scientifique

économique : production « industrielle » de grandes quantités de textes simplifiés

- actuellement, production manuelle « artisanale » et à petite échelle



PROJET « SIMPLES » : PARTENAIRES



- Chef de file
- leader mondial de la **traduction automatique**
- Etablissement Public National Antoine Koenigswarter
- accompagne près de 3 200 personnes en situation de **handicap intellectuel et psychique**
- Laboratoire d'Etude des Usages en Technologies d'Information Numériques
- **usages des nouvelles technologies** dans le cadre de recherches académiques et industrielles



Holken Consultants
& Partners

Holken Consultants en sous-traitance pour l'étude "Recueil des besoins"



PROJET « SIMPLES » : SIMPLIFICATION ET FALC

FALC - « facile à lire et à comprendre »

- méthode élaborée dans le cadre du projet européen Pathways
- portée en France par l'Unapei et l'association Nous Aussi  **Unapei**
POUR UNE SOCIÉTÉ SOLIDAIRE ET INCLUSIVE
- règles pour rédiger les documents pour les personnes déficientes intellectuelles
 - + 20 règles générales et 42 règles spécifiques aux informations écrites
 - + règles prioritaires, règles importantes, règles optionnelles...
 - + règles de construction, règles de mise en page, illustrations...
- certification, logo 

BARRIERES

- règles trop exigeantes et trop nombreuses
- mise en page contraignante
- exclut les autres populations



PROJET « SIMPLES » : ETAPES

ETUDES PREALABLES

Etude des pratiques de mise en FALC manuelle

- Formation FALC pour les participants du projet
- Interviews des personnels des ESAT
- Formalisation du savoir-faire du FALC
- Spécifications pour l'architecture de la plateforme Open Source d'aide à la transcription en FALC



Etude de recueil des besoins en simplification et en FALC

- Recensement des pratiques existantes de la simplification et la mise en FALC des documents
- Grand intérêt et une véritable nécessité d'adopter la démarche de simplification et d'accessibilité
- Difficultés d'automatisation



PROJET « SIMPLES » : ETAPES

DEVELOPPEMENT DES OUTILS

“LIREC” : plateforme Open Source d’aide à la mise en FALC

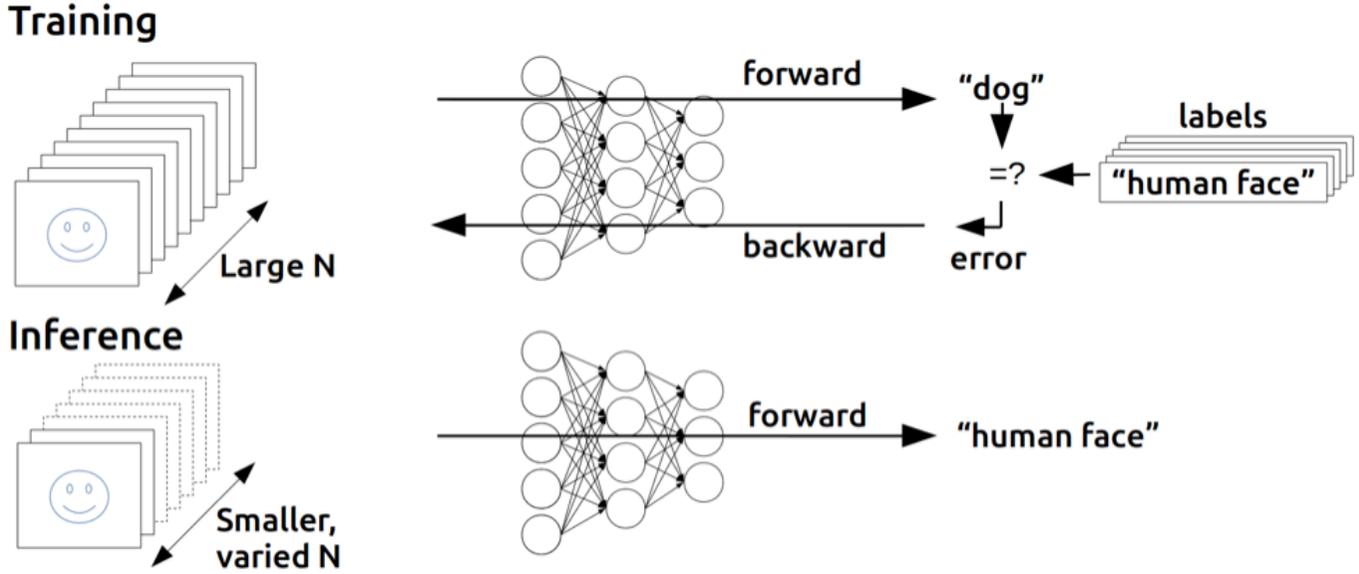
- Suggestions de mise en forme FALC
- Suggestions de simplification des mots
- Recherche de pictogrammes
- Estimation de la complexité globale du texte
- Proposition de modèles de « *frames* » prédéfinis selon la thématique du texte



PROJET « SIMPLES » : ETAPES

DEVELOPPEMENT DES OUTILS

Simplification Automatique : outil basé sur les technologies neuronales

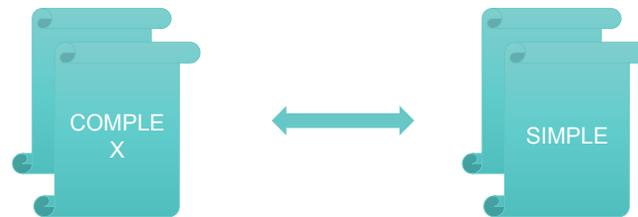


PROJET « SIMPLES » : ETAPES

DEVELOPPEMENT DES OUTILS

Simplification Automatique : difficultés

- Manque de corpus aligné pour l'apprentissage « supervisé »



- **ANGLAIS**
corpus peu nombreux et bruités



- **FRANCAIS**
presque rien...

- Tâche difficile à définir Une « bonne » simplification, c'est quoi ?
 - peut être plus courte (résumé) ou plus longue (explication)
 - dépend du public visé
 - accord entre-annotateurs faible



PROJET « SIMPLES » : ETAPES

DEVELOPPEMENT DES OUTILS

Simplification Automatique : approches expérimentales, **approximations**

1. Simplification = Résumé

- texte plus court = texte plus lisible
- corpus collectés à partir de **résumés d'actualités**

2. Techniques « non-supervisées »

- Corpus non-alignés disponibles en abondance
- Qualité plus faible que pour les systèmes supervisés

3. Traduction automatique Systran des corpus anglais vers le français

- EN -> FR
- Risque de bruit supplémentaire



Avez-vous des questions ?



Merci !

Natalia Segal, Systran Group
natalia.segal@systrangroup.com