



Linguistique et exigences

Anne Condamines, Maxime Warnier, laboratoire CLLE (Cognition, Langues, Langage, Ergonomie)^{1,2}, Toulouse

1- CNRS

2-Université Jean Jaurès

Problématique

- Dans les entreprises, la langue est utilisée comme moyen de transmission de l'information
- Or, le rôle de la langue n'est pas seulement d'encoder de l'information
- La polysémie et l'imprécision demeurent même dans les situations professionnelles

Exemple : le chef de projet est responsable de la documentation du projet (processus vs résultat)

Tentatives pour pallier ces difficultés : la création de langues contrôlées (CNL)

- *Recommandations linguistiques visant à améliorer la qualité des échanges en milieu professionnel*
- Apparues pour la première fois chez Caterpillar au début des années 1970
- Problème : l'efficacité de ces Controlled Natural Languages (CNL) a été très peu évaluée
- Or, les CNL sont parfois trop contraignants (*ne pas utiliser de conjonctions*) et parfois pas assez (problème des nominalisations).

Proposition pour améliorer la constitution des CNL : Mettre en place une « linguistique ergonomique (LE) »

- La LE utilise l'expérience de l'ergonomie, en particulier la notion d'*utilisabilité*
Degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficacité et satisfaction dans un contexte d'utilisation spécifié (Iso 9241-11)
- La LE s'inspire des usages réels
 - Pour proposer des normes qui soient proches de la réalité
 - Pour repérer des régularités qui se mettent spontanément en place dans les groupes de locuteurs
- Les propositions faites pour constituer un CNL sont évaluées
 - Par des tests d'acceptabilité auprès des locuteurs concernés
 - Par des expérimentations psycholinguistiques (thèse avec Airbus)

Rédaction des exigences au CNES

- ▶ Thèse M. Warnier : « Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique : l'exemple des exigences de projets spatiaux »
- ▶ Objectif : proposer une méthode de constitution d'une langue contrôlée pour la rédaction des exigences au CNES
 - ▶ limitant au maximum les risques langagiers
 - ▶ aussi naturelle que possible
- ▶ Étude sur corpus

Méthodologie

- ▶ Au CNES, pas de LC imposée
- ▶ Hypothèse : malgré tout des pratiques récurrentes (dus à l'expérience des rédacteurs, leurs intuitions, l'influence indirecte des LC existantes, etc.)
 - ▶ Un « genre » des exigences
- ▶ Proposition : se servir de ces régularités pour proposer des règles de rédaction
- ▶ Deux façons de procéder :
 1. Vérifier si certaines régularités supposées se constatent effectivement
 2. Faire émerger les régularités du corpus

Application : méthode 1

- LC analysée (en anglais) : *Guide for Writing Requirements* d'INCOSE (2011)
 - Pour la rédaction des exigences
 - Inspiré de normes antérieures
- “Avoid combinator”
 - “Combinators are words that join clauses together, such as 'and', 'or', 'then', 'unless'. Their presence in a requirement usually indicates that multiple requirements should be written.”
- “Repeat nouns in full instead of using pronouns to refer to nouns in other requirement statements”
 - “Pronouns are words such as 'it', 'this', 'that', 'he', 'she', 'they', 'them'. When writing stories, they (sic.) are a useful device for avoiding the repetition of words; but when writing requirements, pronouns should be avoided, and the proper nouns repeated where necessary.”

Méthode 1 : analyse quantitative

	Pronoms	Conjonctions coordination	Conjonctions subordination	Conjonctions (total)
Exigences	1,26 %	2,21 %	0,64 %	2,85 %
Cours spatial	2,59 %	2,69 %	0,87 %	3,56 %
Presse	3,88 %	2,44 %	1,30 %	3,75 %

Méthode 1 : analyse qualitative

► Conjonctions :

► Indispensables

- « Le générateur de TCH vérifiera **que** la valeur du champ PHASE est comprise entre 0 **et** `FREQ_DIV -1`. »

► Justifiées (?)

- « Ces contrôles seront implémentés dans GOTLIB **mais** devront être explicitement appelés par le DUPC »
- « Les champs `SM_ID` **et** `FM_ID` seront extraits à partir de la BDS »

► Indésirables

- « Le format des données de mesure angulaire et Doppler est conforme au standard CCSDS décrit dans le document DA9 **et** le schéma XML respecte le standard décrit dans DA11.»

Méthode 1 : analyse qualitative

► Pronoms :

► Indispensables

- « Sur réception de cette TC, le LVC met à jour le paramètre **qui** donne la taille maximum d'un paquet TM de type dump »

► Justifiés (?)

- « Le paquet ne sera généré que **s'il** est activé par le LVC. »
- « Le générateur de TC ne rejettera pas la création du PARAM_ID diagnostic si **celui-ci** est déjà défini à bord. »

► Indésirables

- « **Il** calculera aussi, a une fréquence paramétrable (ordre de grandeur 1 mois), la moyenne de mise en œuvre »

Méthode 1 : analyse qualitative

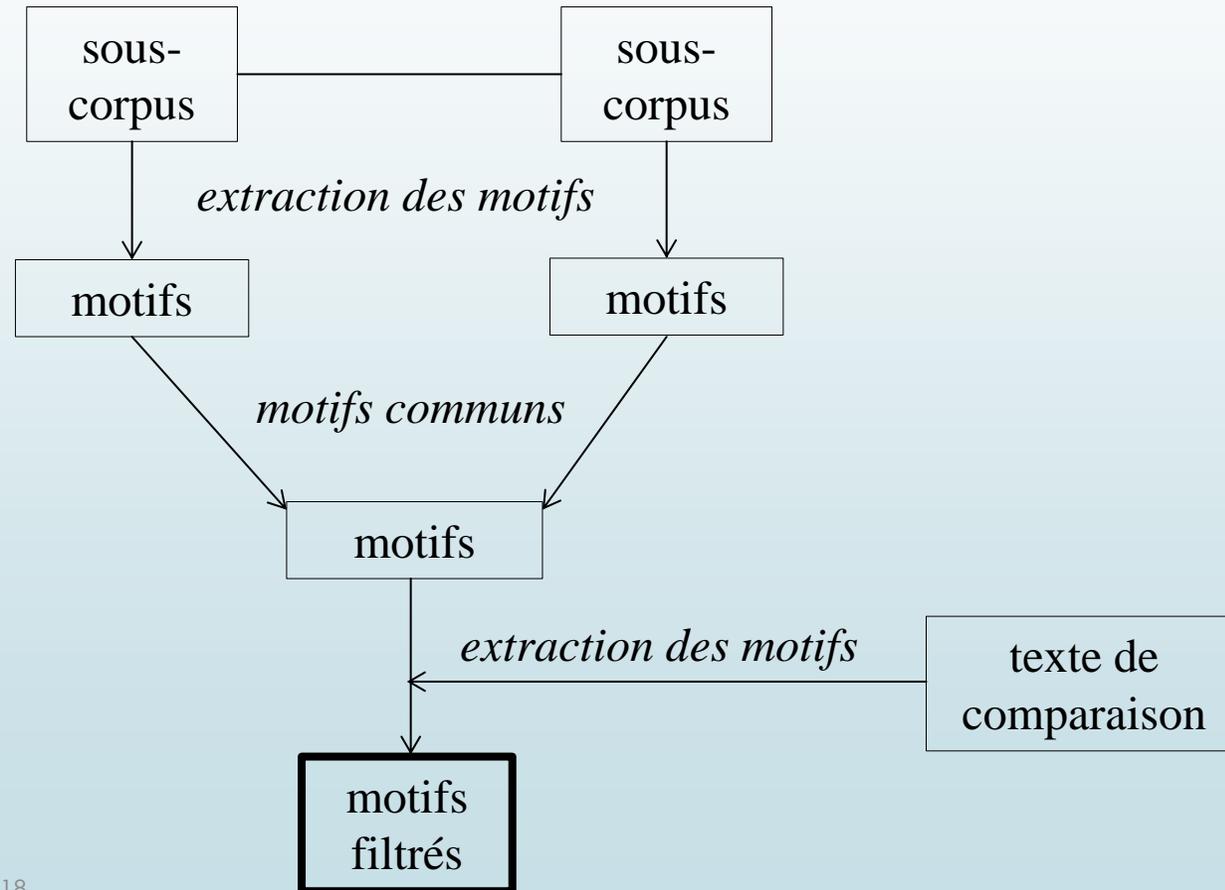
► Pronoms :

- Certains pronoms problématiques
- Pas tous les pronoms / pas que les pronoms
- Règle plus efficace : « N'utilisez un pronom que s'il a un et un seul antécédent possible au sein de l'exigence »
- Aide à la rédaction : signaler les pronoms qui ont plusieurs antécédents possibles

Application : méthode 2

- ▶ Extraction de motifs séquentiels (SDMC)
- ▶ *Boilerplates*
 - ▶ ex. : « Quand [condition], le [système] doit [action] »
 - ▶ Simples et utiles, mais souvent trop génériques

Application : méthode 2



Méthode 2 : résultats

- ▶ « La durée maximale/maximum [avant délivrance] est de [10 minutes] »
- ▶ « La durée [des activités] est limitée à vs ne doit pas dépasser vs ne doit pas excéder vs doit être inférieure ou égale à [2 jours] »
- ▶ *Le/la <propriété> <complément> doit être <inférieur(e) | supérieur(e)> [ou égal(e)] à <nombre> <unité>*

Évaluation des résultats

- Test d'acceptabilité : « quelle(s) structure(s) jugez-vous préférable(s) ? » (note sur 5 pour chaque proposition)

	1	2	3	4	5
L'utilisateur doit savoir à tout moment s'il est connecté sur le serveur GIDE nominal ou redondant.	<input type="radio"/>				
L'utilisateur doit savoir à tout moment si cet utilisateur est connecté sur le serveur GIDE nominal ou redondant.	<input type="radio"/>				
L'utilisateur doit savoir à tout moment si l'utilisateur est connecté sur le serveur GIDE nominal ou redondant.	<input type="radio"/>				

Évaluation des résultats

- ▶ L'utilisateur doit savoir à tout moment s'il est connecté sur le serveur GIDE nominal ou redondant. **[4,95]**
- ▶ L'utilisateur doit savoir à tout moment si l'utilisateur est connecté sur le serveur GIDE nominal ou redondant. **[2,01]**
- ▶ L'utilisateur doit savoir à tout moment si cet utilisateur est connecté sur le serveur GIDE nominal ou redondant. **[2,00]**

Conclusion

- Objectif : proposer une méthode outillée réutilisable de constitution/amélioration des règles de rédaction des exigences adaptées aux pratiques réelles (*utilisabilité*)
 - Analyses de données attestées
 - Recherche de régularités linguistiques
 - Validation des hypothèses
 - Évaluation des résultats par les utilisateurs
- Perspectives :
 - Implémentation des règles
 - Évaluation de la compréhension réelle
 - Automatisation des étapes

Merci !