

Identifier et catégoriser l'ambiguïté dans les spécifications techniques de conceptions de systèmes

Émilie Merdy

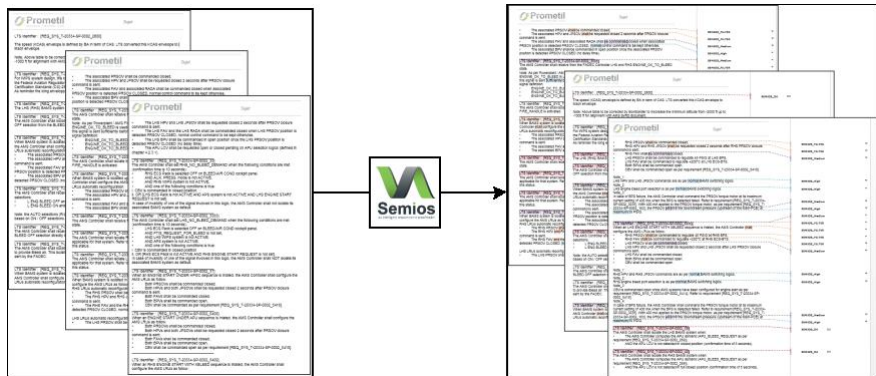
CLLE-ERSS / Prometil

RÉCITAL 4-8 juillet 2016

Thèse débutée en octobre 2015 (Cifre)



Semios : détection des erreurs dans les documents techniques



Kang and Saint-Dizier (2015), Kang and Merdy (2015)

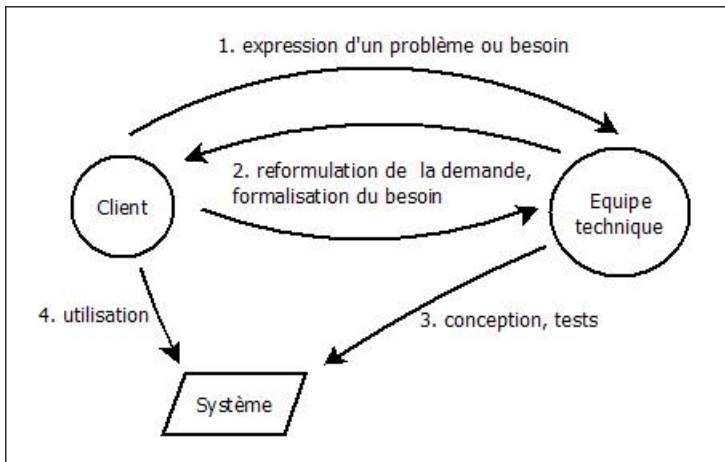
- 1 Contexte : besoin et définitions
- 2 Caractéristiques du corpus
- 3 Premiers éléments de résolution
- 4 Bilan et perspectives

- 1 Contexte : besoin et définitions
- 2 Caractéristiques du corpus
- 3 Premiers éléments de résolution
- 4 Bilan et perspectives

Spécifications

- 3^{ème} critère de réussite d'un projet (The Standish Group (2013))
- cause (partielle) de 50% des échecs (Hastie and Wojewoda (2015))
- 72% rédigées en langage naturel (Berry et al. (2003))

Succès : atteinte d'un objectif sans dépassement **critique** d'un délai et/ou d'un budget imparti



Cycle de vie simplifié d'une spécification

Spécification - *specification*

Spécification - *specification*

Du texte

ID_12

The ARTS shall provide voltage information based on fluid air temperature measured.

[...]

ID_24

The main HSVP shall compute APU temperatures every 7 minutes.

Spécification - *specification*

Du texte

ID_12

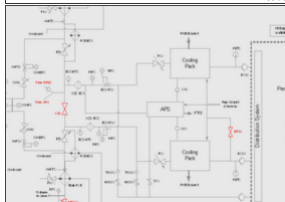
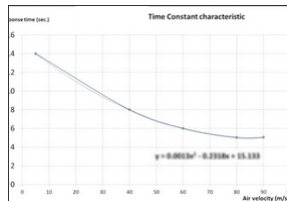
The ARTS shall provide voltage information based on fluid air temperature measured.

[...]

ID_24

The main HSVP shall compute APU temperatures every 7 minutes.

+ figures, formules ...



Spécification - *specification*

Document technique à valeur contractuelle

Du texte

ID_12

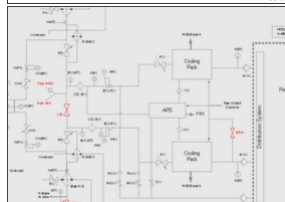
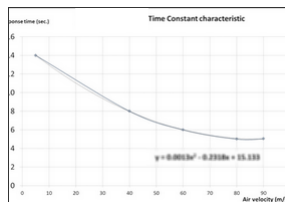
The ARTS shall provide voltage information based on fluid air temperature measured.

[...]

ID_24

The main HSVP shall compute APU temperatures every 7 minutes.

+ figures, formules ...



Exigence - *requirement*

Exemple

ID_12

The ARTS shall provide voltage information based on fluid air temperature measured.

Exigence - *requirement*

*“Une déclaration qui identifie une caractéristique ou une contrainte opérationnelle, fonctionnelle ou de conception relative à un produit ou à une procédure, qui est **non-ambiguë**, testable ou mesurable, et nécessaire à l’acceptabilité du produit ou de la procédure” IEEE (1998)*

Exemple

ID_12

The ARTS shall provide voltage information based on fluid air temperature measured.

Une *bonne* exigence doit être :

- **non-ambiguë**, simple, concise, courte, lisible, testable, mesurable, vérifiable, complète, atomique, unique, identifiée, traçable, nécessaire, justifiée, etc.

(IEEE/ISO/IEC (2011), INCOSE, IREB et guides de rédaction internes)

Une *bonne* exigence doit être :

- **non-ambiguë**, simple, concise, courte, lisible, testable, mesurable, vérifiable, complète, atomique, unique, identifiée, traçable, nécessaire, justifiée, etc.

(IEEE/ISO/IEC (2011), INCOSE, IREB et guides de rédaction internes)

Une bonne exigence ne doit pas :

- être imprécise, être trop longue, contenir plus d'une attente, dépendre d'une autre exigence, contenir un autre modal que "shall", etc.

De nombreux travaux antérieurs concernant :

- l'ambiguïté : Shapiro and Snyder (2015), Zhang (1998),
- l'ambiguïté et les contextes spécialisés/techniques : Kuhn (2014), Condamines and Rebeyrolle (1996),
- l'ambiguïté/l'intelligibilité dans les exigences : Condamines and Warnier (2014), Berry et al. (2012), Tjong (2008).

Enjeux

Enjeux

1/ **Ambiguïté perçue :**

- incertitudes chez le
lecteur-utilisateur

⇒ prise de décision ou impasse

Enjeux

1/ Ambiguïté perçue :

- incertitudes chez le
lecteur-utilisateur

⇒ prise de décision ou impasse

2/ Ambiguïté non perçue :

- choix pour une interprétation
différente ?

⇒ risque de problème non détecté

Enjeux

1/ Ambiguïté perçue :

- incertitudes chez le
lecteur-utilisateur

⇒ prise de décision ou impasse

2/ Ambiguïté non perçue :

- choix pour une interprétation
différente ?

⇒ risque de problème non détecté

Objectif

⇒ Mise au point d'une méthodologie robuste de création
semi-automatique de ressources linguistiques adaptées aux domaines

- 1 Contexte : besoin et définitions
- 2 **Caractéristiques du corpus**
- 3 Premiers éléments de résolution
- 4 Bilan et perspectives

Corpus d'exigences

- 5 000 exigences issues de 5 spécifications,
- Anglais technique
- Variations lexicale et syntaxique faibles (4 500 types/220 000 tokens)

Corpus d'exigences

- 5 000 exigences issues de 5 spécifications,
- Anglais technique
- Variations lexicale et syntaxique faibles (4 500 types/220 000 tokens)

Caractéristiques des utilisateurs des spécifications

- nombreux
- différents postes
- langue maternelle
- formés à différents standards

⇒ correcteurs attendent Précision > Rappel

Type	Exemples
Approximation	<i>The adjacent TCV estimated position shall be acquired from adjacent channel by XTalk every 50 ms if valid, set to its last valid value otherwise.</i>
Portée syntaxique	
Polysémie	
Anaphore	
Sous-spécification	

Type	Exemples
Approximation	<i>Over current protection circuit shall be activated at min 500 mA whatever the duration [...]</i>
Portée syntaxique	
Polysémie	
Anaphore	
Sous-spécification	

Type	Exemples
Approximation	<i>The system shall not suffer any deformation and operates normally after the application of clamp assembly [...]</i>
Portée syntaxique	
Polysémie	
Anaphore	
Sous-spécification	

Cas actuellement détectés par Semios

Type	Traité	Comment
Approximation	Oui	Lexique générique + règles morpho-syntaxiques
Portée syntaxique	Oui	Lexique générique + règles morpho-syntaxiques
Polysémie	Non	
Anaphore	Partiellement	Lexique générique + règles morpho-syntaxiques
Sous-spécification	Partiellement	Lexiques spécifiques

- 1 Contexte : besoin et définitions
- 2 Caractéristiques du corpus
- 3 Premiers éléments de résolution**
- 4 Bilan et perspectives

Détecter l'ambiguïté lexicale simple : méthodologie

Détecer l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

Détecter l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

“about, almost, around, minimum...”

Déteçter l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

“about, almost, around, minimum...”

2/ Filtres morpho-syntaxiques

Détecter l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

“about, almost, around, minimum...”

2/ Filtres morpho-syntaxiques

*“The **maximum** pressure loads [...] shall be 6.”*

Détecter l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

“about, almost, around, minimum...”

2/ Filtres morpho-syntaxiques

*“The **maximum** pressure loads [...] shall be 6.”*

3/ Signaler les termes flous

Détecter l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

"about, almost, around, minimum..."

2/ Filtres morpho-syntaxiques

*"The **maximum** pressure loads [...] shall be 6."*

3/ Signaler les termes flous

*"[...] shall be activated at **min 500 mA**"*

Détecter l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

"about, almost, around, minimum..."

2/ Filtres morpho-syntaxiques

*"The **maximum** pressure loads [...] shall be 6."*

3/ Signaler les termes flous

*"[...] shall be activated at **min 500 mA**"*

Idem pour la reprise anaphorique, la portée de la négation...

Détecter l'ambiguïté lexicale simple : méthodologie

1/ Lexique des quantifieurs flous

"about, almost, around, minimum..."

2/ Filtres morpho-syntaxiques

*"The **maximum** pressure loads [...] shall be 6."*

3/ Signaler les termes flous

*"[...] shall be activated at **min 500 mA**"*

Idem pour la reprise anaphorique, la portée de la négation...

Lexiques génériques

⇒ construction manuelle OK

Sous-spécifié

The **system** shall deliver...

The **interface** shall deliver...

The **XX interface** shall deliver...

Spécifié

Hyperonyme (générique)



Hyponyme (générique)



Hyponyme (spécifié)

Déteçter l'ambiguïté des termes génériques sous-spécifiés

Détecter l'ambiguïté des termes génériques sous-spécifiés

1/ Lexique de termes génériques

Détecer l'ambiguïté des termes génériques sous-spécifiés

1/ Lexique de termes génériques

“system, component...” / “interface, engine...”

Détecer l'ambiguïté des termes génériques sous-spécifiés

1/ Lexique de termes génériques

“system, component...” / “interface, engine...”

⇒ couverture lexicale + : synonymes, thésaurus, voisins distributionnels

(Merdy et al. (2016))

Détecter l'ambiguïté des termes génériques sous-spécifiés

1/ Lexique de termes génériques

“system, component...” / “interface, engine...”

⇒ couverture lexicale + : synonymes, thésaurus, voisins distributionnels

(Merdy et al. (2016))

2/ Lexique des termes spécifiés pour limiter les faux-positifs

Détecer l'ambiguïté des termes génériques sous-spécifiés

1/ Lexique de termes génériques

“system, component...” / “interface, engine...”

⇒ couverture lexicale + : synonymes, thésaurus, voisins distributionnels

(Merdy et al. (2016))

2/ Lexique des termes spécifiés pour limiter les faux-positifs

“the RCS system” / “the XX interface”

Détecer l'ambiguïté des termes génériques sous-spécifiés

1/ Lexique de termes génériques

“system, component...” / “interface, engine...”

⇒ couverture lexicale + : synonymes, thésaurus, voisins distributionnels

(Merdy et al. (2016))

2/ Lexique des termes spécifiés pour limiter les faux-positifs

“the RCS system” / “the XX interface”

3/ Signalement des termes génériques lorsqu'ils sont sous-spécifiés

- 1 Contexte : besoin et définitions
- 2 Caractéristiques du corpus
- 3 Premiers éléments de résolution
- 4 Bilan et perspectives**

Travaux en cours : bilan et perspectives

Travaux en cours : bilan et perspectives

- Sous-catégorisation de l'ambiguïté lexicale

Travaux en cours : bilan et perspectives

- Sous-catégorisation de l'ambiguïté lexicale
- Niveaux de généralité/spécificité

Travaux en cours : bilan et perspectives

- Sous-catégorisation de l'ambiguïté lexicale
- Niveaux de généralité/spécificité
- Définition des *frontières* du domaine

Travaux en cours : bilan et perspectives

- Sous-catégorisation de l'ambiguïté lexicale
- Niveaux de généralité/spécificité
- Définition des *frontières* du domaine
- Construction de corpus spécifiques (Baroni and Bernardini (2004), Duthil et al. (2011))

Travaux en cours : bilan et perspectives

- Sous-catégorisation de l'ambiguïté lexicale
- Niveaux de généralité/spécificité
- Définition des *frontières* du domaine
- Construction de corpus spécifiques (Baroni and Bernardini (2004), Duthil et al. (2011))
- Application de l'Analyse Distributionnelle Automatique pour construire des lexiques de termes génériques

- Baroni, M. and Bernardini, S. (2004). Bootcat : Bootstrapping corpora and terms from the web. In *LREC*.
- Berry, D., Gacitua, R., Sawyer, P., and Tjong, S. F. (2012). The case for dumb requirements engineering tools. In *Requirements engineering : Foundation for software quality*, pages 211–217. Springer.
- Berry, D. M., Kamsties, E., and Krieger, M. M. (2003). From contract drafting to software specification : Linguistic sources of ambiguity. Technical report, Technical Report, School of Computer Science, University of Waterloo, Waterloo, ON, Canada.
- Condamines, A. and Rebeyrolle, J. (1996). Point de vue en langue spécialisée. *Meta*, 42(1) :174–184.
- Condamines, A. and Warnier, M. (2014). Linguistic analysis of requirements of a space project and their conformity with the recommendations proposed by a controlled natural language. In *International Workshop on Controlled Natural Language*, pages 33–43. Springer.
- Duthil, B., Troussel, F., Roche, M., Dray, G., Plantié, M., Montmain, J., and Poncelet, P. (2011). Towards an automatic characterization of criteria. In *Database and Expert Systems Applications*, pages 457–465. Springer.
- Hastie, S. and Wojewoda, S. (2015). Standish group 2015 chaos report-q&a with jennifer lynch. *Retrieved*, 1(15) :2016.
- IEEE (1998). IEEE 1220-1994. Technical report, Rapport technique.
- IEEE/ISO/IEC (2011). IEEE, ISO, IEC. 29148 : 2011-systems and software engineering-requirements engineering. Technical report, Rapport technique.
- Kang, J. and Merdy, E. (2015). Semios : Relecteur automatique d'exigences pour une aide à la rédaction de spécifications. *Génie Logiciel*, 115 :10–16.
- Kang, J. and Saint-Dizier, P. (2015). Une expérience d'un déploiement industriel de liele : une relecture intelligente des exigences. In *Actes de INFORSID*.
- Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, 40(1) :121–170.
- Merdy, E., Kang, J., and Tanguy, L. (2016). Identification de termes flous et génériques dans la documentation technique : expérimentation avec l'analyse distributionnelle automatique. In *23e Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2016)*.
- Shapiro, S. and Snyder, E. (2015). Vagueness and context. *Inquiry*, pages 1–39.
- The Standish Group (2013). Chaos manifesto.
- Tjong, S. F. (2008). *Avoiding ambiguity in requirements specifications*. PhD thesis, Citeseer.
- Zhang, Q. (1998). Fuzziness-vagueness-generality-ambiguity. *Journal of pragmatics*, 29(1) :13–31.

Merci de votre attention !
Des questions ?

*Je pense que la fraternité,
c'est un état de confusion grammaticale
entre je et eux, moi et lui, avec possibilités.*

É. Ajar