

TextCoop - LELIE :
Une boîte à outils pour l'analyse des exigences

Patrick Saint-Dizier
IRIT – CNRS Toulouse

stdizier@irit.fr



Défis et applications

La plupart des specs et des exigences sont écrites en langue naturelle:

- Cela soulève de nombreuses difficultés de modélisation (ex. langue linéaire, floue, implicite)
- Cela entraîne de nombreux types d'erreurs ou d'imprécisions qu'il faut maîtriser
- Développer un ensemble d'outils modulaires et adaptables aux situations et aux rédacteurs d'exigences.

Sommaire de la présentation:

- Problèmes et défis de la rédaction d'exigences en langue naturelle
- TextCoop et LELIE: 2 projets de recherche ANR
- La plateforme TextCoop pour l'analyse du discours technique
- LELIE: un assistant pour écrire des exigences selon vos normes
- Tâches en cours et Projets

Améliorer la qualité des exigences:

(1) les défis de la rédaction en langue naturelle

→ **Comment s'assurer qu'un ensemble d'exigences est:**

- suffisamment clair,
- minimalement ambigu,
- et complet par rapport au point à traiter?

But: éviter des incompréhensions et erreurs qui entraînent:

- *Des erreurs de conception et des caractéristiques inadéquates,*
- *Un manque de traçabilité,*
- *Des erreurs en production et des coûts plus élevés (+50%),*
- *Des accidents humains, écologiques, sociaux, etc.*

La langue naturelle:

(2) Challenges sur l'acceptabilité et la cohérence

→ Comment s'assurer que des exigences sont acceptables par les lecteurs et opérateurs ? Elles doivent être:

- *comprises,*
- *acceptées,*
- *faisables et*
- *sans considérations inutiles.*

→ Comment s'assurer qu'un ensemble d'exigences a:

- une bonne **cohérence** interne
- une bonne **cohésion** textuelle ?

→ **Contrôles de qualité sur la forme et le contenu,**

→ **Mesures de qualité.**

Notre solution: TextCoop et LELIE

→ **TextCoop**: analyse de la structure d'un texte (discours):

- passage langue naturelle vers représentation semi-formelle.

→ **LELIE: boîte à outils d'aide à la rédaction d'exigences**, détecte des erreurs typiques non traitées par les éditeurs de textes standard:

- (1) suit les principes et normes de l'écriture des exigences,
- (2) produit des alertes lorsque des erreurs sont détectées et suggère des corrections,
- (3) couvre une grande diversité de types d'exigences,
- (4) peut être adapté à divers domaines et contraintes d'entreprises.

Disponibilité du code:

→ **TextCoop V5.1 + LELIE V2.1: Code Open Source.**

Position de TextCoop + LELIE

- **TextCoop:**

Très peu de systèmes d'analyse de discours dans le monde.

Utilisé par environ 25 sites industriels et d'enseignement en Europe.

- **LELIE:**

systèmes proches sur le marché:

- principal: **RAT-RQA** Reusecompany, basé sur des gabarits + contrôles cohésion et cohérence, philosophie très différente de LELIE.

- quelques autres, plus limités: **Lexior** (Cortim), **error checker** (Synapse), **Rubric** (univ. of Luxembourg), **Captilo** (Prolipsia), etc.

→ **Prototype: stable, testé et évalué sur des cas réels (env. TRL 5).**

Sommaire de la présentation:

- Problèmes et défis en rédaction d'exigences en langue naturelle
- TextCoop et LELIE: 2 projet de recherche ANR

→ La plateforme TextCoop pour l'analyse du discours technique

- LELIE: votre assistant pour écrire des exigences selon les normes
- Tâches en cours et Projets

Partie 1: La Plateforme <TextCoop> dédiée à l'analyse du discours

→ Reconnaît les structures suivantes:

- langue générale:

conditions, circonstances, justifications, illustrations, reformulations, élaborations, buts, conclusions, sommaires, etc.

- spécifiques aux documents techniques:

exigences, instructions, titres, prérequis, définitions, avertissements, conseils, etc.

→ Niveau de la phrase:

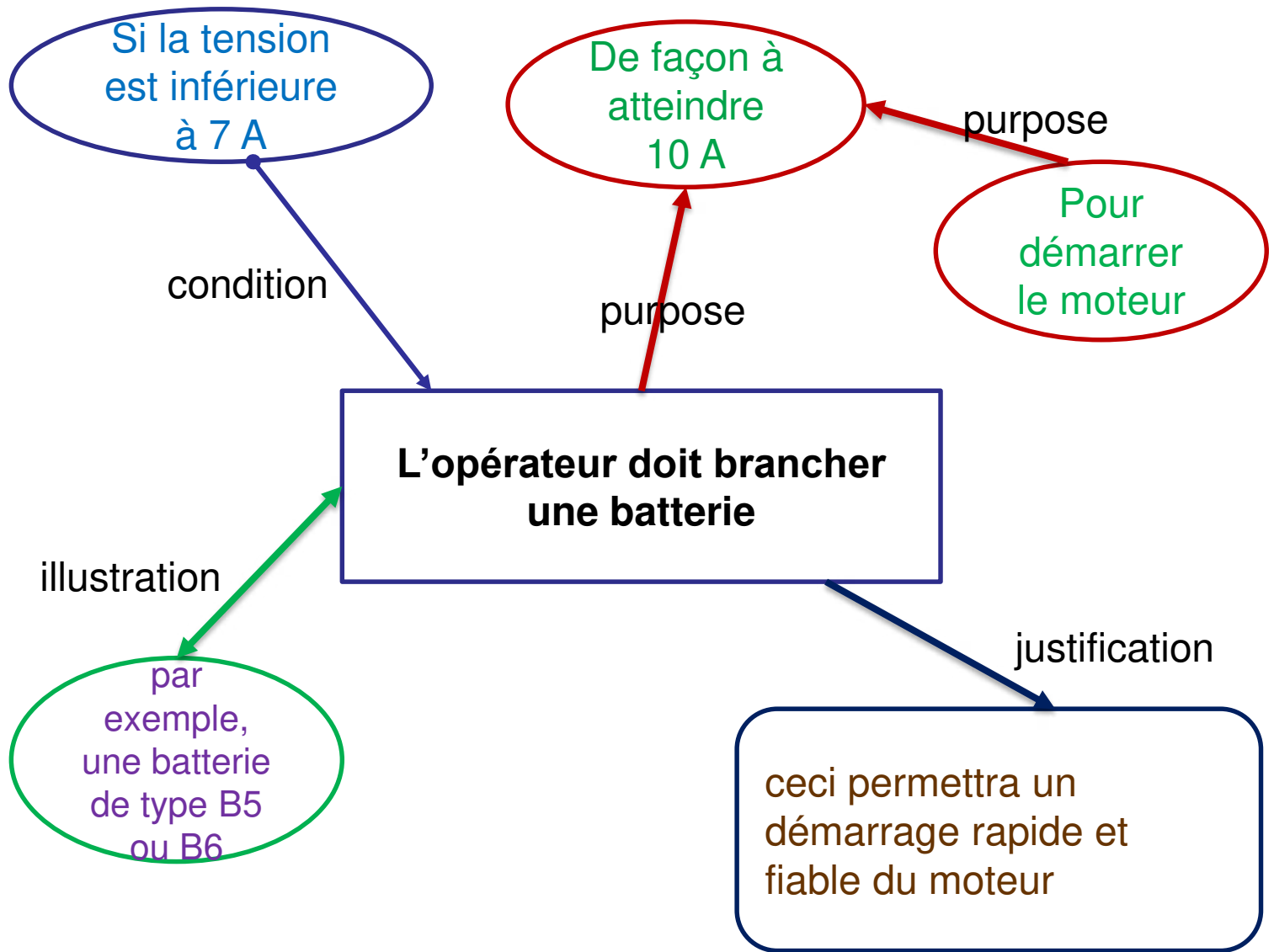
- structures typiques: verbes et compléments, expressions temporelles et spatiales, expressions du moyen et de l'instrument, etc.

Une exigence avec ses structures de discours

[[condition Si la tension est inférieure à 7 A],
[purpose de façon à atteindre 10 A
[purpose pour démarrer le moteur]],

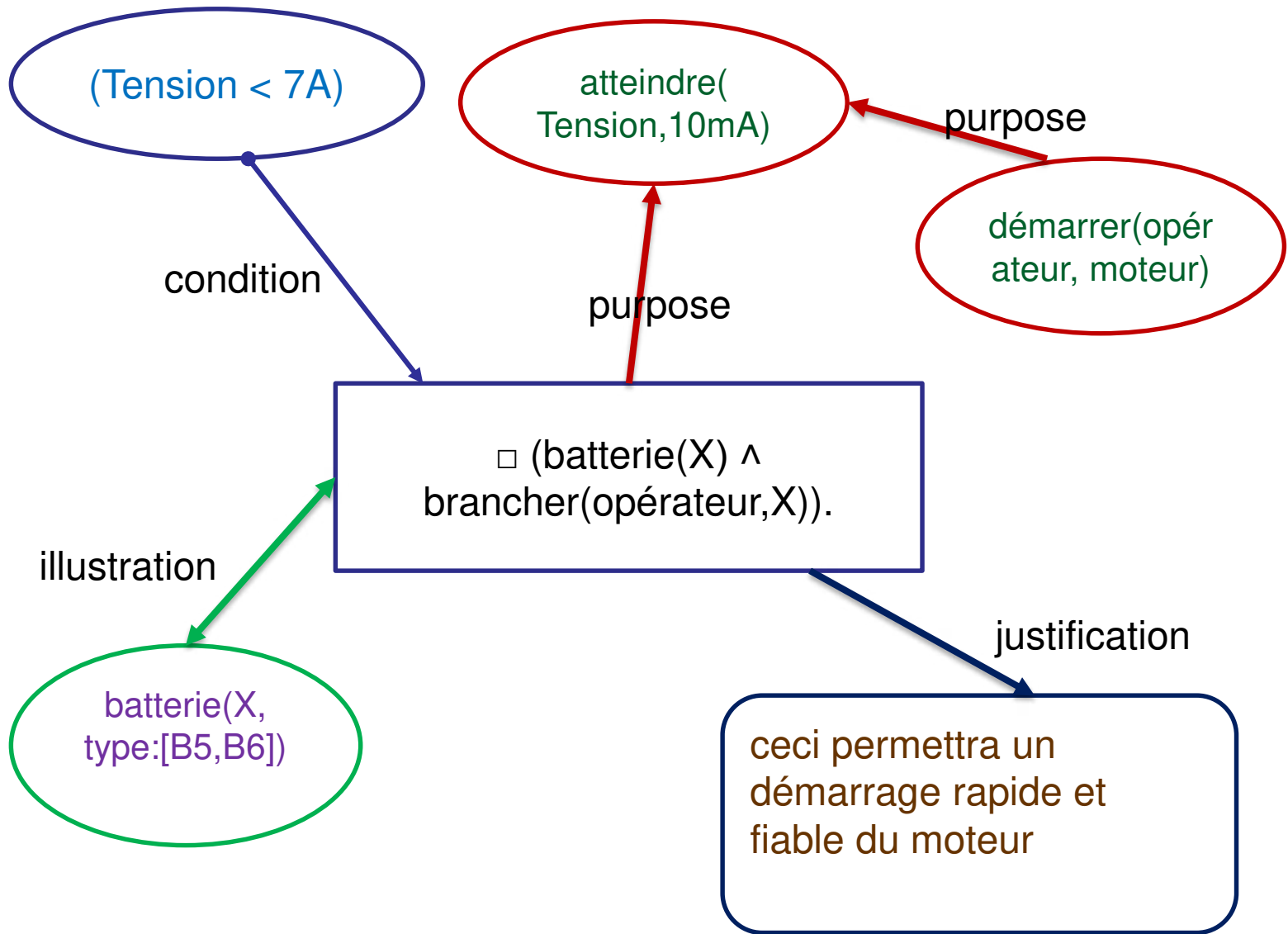
l'opérateur doit brancher une batterie, ← élément central

[illustration par exemple, une batterie de type B5 or B6],
[justification ceci permettra un démarrage rapide et fiable
du moteur.]]



Du langage naturel à une représentation semi-formelle

- Les différentes parties d'une exigence peuvent être représentées plus formellement en logique ou par un système attribut-valeurs typé :
 - La tension est inférieure à 7 A: (Tension < 7A)
 - Une batterie de type B5 ou B6: batterie(X, type:[B5,B6])
 - L'opérateur doit brancher une batterie:
 - (batterie(X) \wedge brancher(opérateur,X)).



Transformer un texte en information structurée avec TextCoop (1)

- **Niveau global: identifie et étiquette (XML):**
 - sommaires, définitions, introduction et préambules,
 - contexte du document,
 - prérequis et équipements éventuels,
 - avertissements et conseils,
 - conclusions, etc.
- **Pour chaque section: identifie et étiquette (XML):**
 - la structure de titres,
 - les introductions, attendus, objectifs, etc.
 - des définitions, contextes et avertissements 'locaux',
 - les exigences et les conditions et justifications attachées
 - les instructions éventuelles (exigences de sécurité).

(2)

- **Pour chaque exigence, développe une représentation partielle semi-formelle:**
 - de l'exigence (exemples ci-dessus),
 - des conditions, contraintes,
 - des justifications

- **Emplois (vers le futur):**
 - *révision, restructuration des documents,*
 - *rédaction collaborative,*
 - *indexation de contenus et*
 - *indicateurs / clusters de traçabilité*

1. Etude du contexte sécurité

Les besoins en sécurité SI d'un Projet sont définis à partir de 4 critères :

- **DISPONIBILITÉ** : les informations doivent être accessibles au moment voulu par les personnes autorisées.

*Un projet a des besoins de disponibilité **lorsqu'il s'inscrit dans le cadre de télé-procédures** (services à disposition des usagers professionnels ou du public depuis Internet, échanges avec d'autres organismes...).*

- **INTÉGRITÉ** : les informations doivent être exactes et complètes.

*Un projet a des besoins d'intégrité **lorsqu'il requiert une ouverture vers des réseaux extérieurs** (transfert de fichiers, accès depuis l'extérieur, accès à des sites externes, ...).*

.....

2. Mise en œuvre opérationnelle des portefeuilles

Considérer le lot 3 ODSDE&A du projet GRDVP. Il requiert l'installation d'une nouvelle définition du portefeuille agent dans AUDE en 13SI2 :

- Suppression de la remise à blanc du référent de suivi principal en cas de changement de modalité d'accompagnement
- Installation du fait générateur du DE dans le portefeuille de l'agent
- Installation du tarif de sortie du DE dans le portefeuille de l'agent
- Mise en place de contrôle de cohérence modalité agent / DE sur les événements de retour en portefeuille

.....

<titre1> 1. Etude du contexte sécurité </titre1>

<préambule> Les besoins en sécurité SI d'un Projet sont définis à partir de 4 critères : </préambule>

- <exigence> **DISPONIBILITÉ** : les informations doivent être accessibles au moment voulu par les personnes autorisées.

<définition> *Un projet a des besoins de disponibilité **lorsqu'il s'inscrit dans le cadre de télé-procédures** <illustration> (services à disposition des usagers professionnels ou du public depuis Internet, échanges avec d'autres organismes...) </illustration>. </définition> </exigence>*

- <exigence> **INTÉGRITÉ** : les informations doivent être exactes et complètes. </exigence>

.....

<titre1> 2. Mise en œuvre opérationnelle des portefeuilles </titre1>

<prérequis> Considérer le lot 3 ODSDE&A du projet GRDVP. </prérequis>

<exigence> Il requiert l'installation d'une nouvelle définition du portefeuille agent dans AUDE en 13SI2 :

- <instruction> Suppression de la remise à blanc du référent de suivi principal en cas de changement de modalité d'accompagnement </instruction>
 - <instruction> Installation du fait générateur du DE dans le portefeuille de l'agent </instruction>
 - <instruction> Installation du tarif du DE dans le portefeuille de l'agent </instruction>
- </exigence>

Objectif:

Transformer un texte fonctionnel en semi-formel

(1)

<exigence> INTÉGRITÉ : les informations doivent être accessibles à tout moment. **</exigence>**

En semi-formel (logique rendue lisible):

Pour tout T, Informations: nécessaire(accessible(Informations, T))

(2)

<instruction> Installation du fait générateur du DE dans le portefeuille de l'agent **</instruction>**

En semi-formel:

Installation(fait_générateur(DE),E), dans(portefeuille(agent), E)

Applications de TextCoop en ingénierie des exigences: résumons

- *Analyse du Discours*

- ✓ Passage d'un texte à sa description structurée,
- ✓ Recherche d'exigences dans de grands volumes de textes,
- ✓ Aide à la révision, la restructuration du texte, amélioration de la rédaction collaborative multi usagers,

- *Analyse du Discours + représentations semi-formelles*

- ✓ améliore la traçabilité,
- ✓ permet une recherche approfondie de doublons ou proches,
- ✓ permet le développement de contrôles de cohérence (via un solveur de contraintes), très complexe!

Sommaire de la présentation:

- Problèmes et défis en rédaction d'exigences en langue naturelle
- TextCoop et LELIE: 2 projet de recherche ANR
- La plateforme TextCoop pour l'analyse du discours technique

→ LELIE: un assistant intelligent pour écrire des exigences selon vos normes

- Tâches en cours et Projets

Partie 2: LELIE: un assistant intelligent pour l'écriture d'exigences

Basé sur:

- **Les normes d'écriture:** REQB, iREB, CMMI-DEV, *IEEE830-1998, INCOSE, ASD-STE 1001, etc.*
- **Des propriétés:** *vérifiable, non ambigu, complet, cohérent*
- **Les pratiques des auteurs d'exigences + normes des sociétés**
- **Le retour d'expérience:** *les normes ne sont pas toujours adaptées !*

Fonctions de Lelie

- ✓ **Laisse le rédacteur écrire comme il le souhaite**, puis analyse
- ✓ **Produit des alertes** sur des expressions incorrectes,
- ✓ Laisse les auteurs faire les corrections mais **suggère des corrections** simples (passifs, termes métiers, structure des phrases, etc.)
- ✓ **Produit une mesure de la qualité de rédaction**, via une métrique, pour chaque type d'erreur,
- ✓ **Mémoire de Correction** : mémorise les corrections faites par les auteurs et les généralise pour les proposer ensuite automatiquement,
- ✓ **Auto-adaptatif**: au domaine et à l'entreprise via de l'apprentissage et de l'induction à partir de corpus.

LELIE: Alertes sur des formes incorrectes (1)

✓ **Présentation générale du document:**

- taille des phrases, des paragraphes à contrôler,
- structure et régularité des énumérations,
- forme, fréquence des titres,
- présence de structures attendues: sommaires, prérequis, avertissements de sécurité, par rapport aux normes, etc.

✓ – **Morphologie:**

- passifs à éviter
- futurs, conditionnels interdits
- autres formes: dérivés des verbes, (*fermeture / fermer*), etc. à éviter

✓ – **Aspects lexicaux:**

- termes flous interdits ou à éviter,
- termes préférés à utiliser,
- termes inappropriés, inutiles, buzz, ambigus, etc. à éviter,
- emploi approprié des termes métier,
- emplois appropriés des verbes, etc.

LELIE: Alertes sur des formes incorrectes (2)

✓ Aspects grammaticaux:

- structure incorrecte d'une exigence,
- négations interdites ou à éviter,
- formes référentielles ambiguës (pronoms, etc.),
- coordinations trop longues (et, ou, et/ou),
- trop de relatives, etc.
- phrases incomplètes.

✓ Style:

- uniformité et bonne position des structures de discours,
- bonne position et emplois du verbe, etc.

✓ Cohesion du texte:

- usages stabilisés des termes (verbes, termes métier, etc.),
- stabilité des structures, etc.

Une sortie typique (XML colorié puis Word commenté)

the user shall indicate if this is an industry **common** **ERROR: fuzzy term** middleware or a vendor specific one .

the user shall indicate if the ww / os (for example ui , ghf or tuio) is **provided** **AVOID: passive construction** by a oem or a third party . in the second case , the tender shall detail the name of the manufacturer : has **never** **ADVICE avoid negation** been easier

the producer shall detail the redundancy and distribution of the different software modules between servers, how **progressively** **ERROR: fuzzy term** a process on one ww element **can** **AVOID: modal** take over from a failing process in another ww element or from a process in a failing ww element **AVOID: sentence too long** .

the producer shall define the best organization to avoid competition between **these** groups **ADVICE: unclear pronoun reference** and the predefined group entities , propose the economic model to be **applied** **AVOID: passive construction** between **them** **ADVICE: unclear pronoun reference** and group international wholesalers , and develop internal partnerships **CHECK: too many relative subordinates** **AVOID: sentence too long** .

RQ1: the user shall indicate if the ww / os (for example ui , ghf or tuio) is **provided** by a oem or a third party . in the second case , the tender shall detail the name of the manufacturer : has **never** been easier

Commenté [s1]: EVITER passif

Commenté [s2]: CONSEIL : pas de négation

RQ2: the producer shall detail the redundancy and distribution of the different software modules between servers, how **progressively** a process on one ww element **can** take over from a failing process in another ww element or from a process in a failing ww element.

Commenté [s3]: ERREUR : terme flou

Commenté [s4]: EVITER Modal

Commenté [s5]: EVITER : phrase trop longue

RQ3: the producer shall define the best organization to avoid competition between **these** groups and the predefined group entities, propose the economic model to be **applied** between **them** and group international wholesalers , and develop internal partnerships.

Commenté [s6]: CONSEIL : référence pronom pas claire

Commenté [s7]: EVITER : passif

Commenté [s8]: CONSEIL : référence pronom pas claire

Commenté [s9]: CONTROLER : trop de relatives

Commenté [s10]: EVITER phrase trop longue

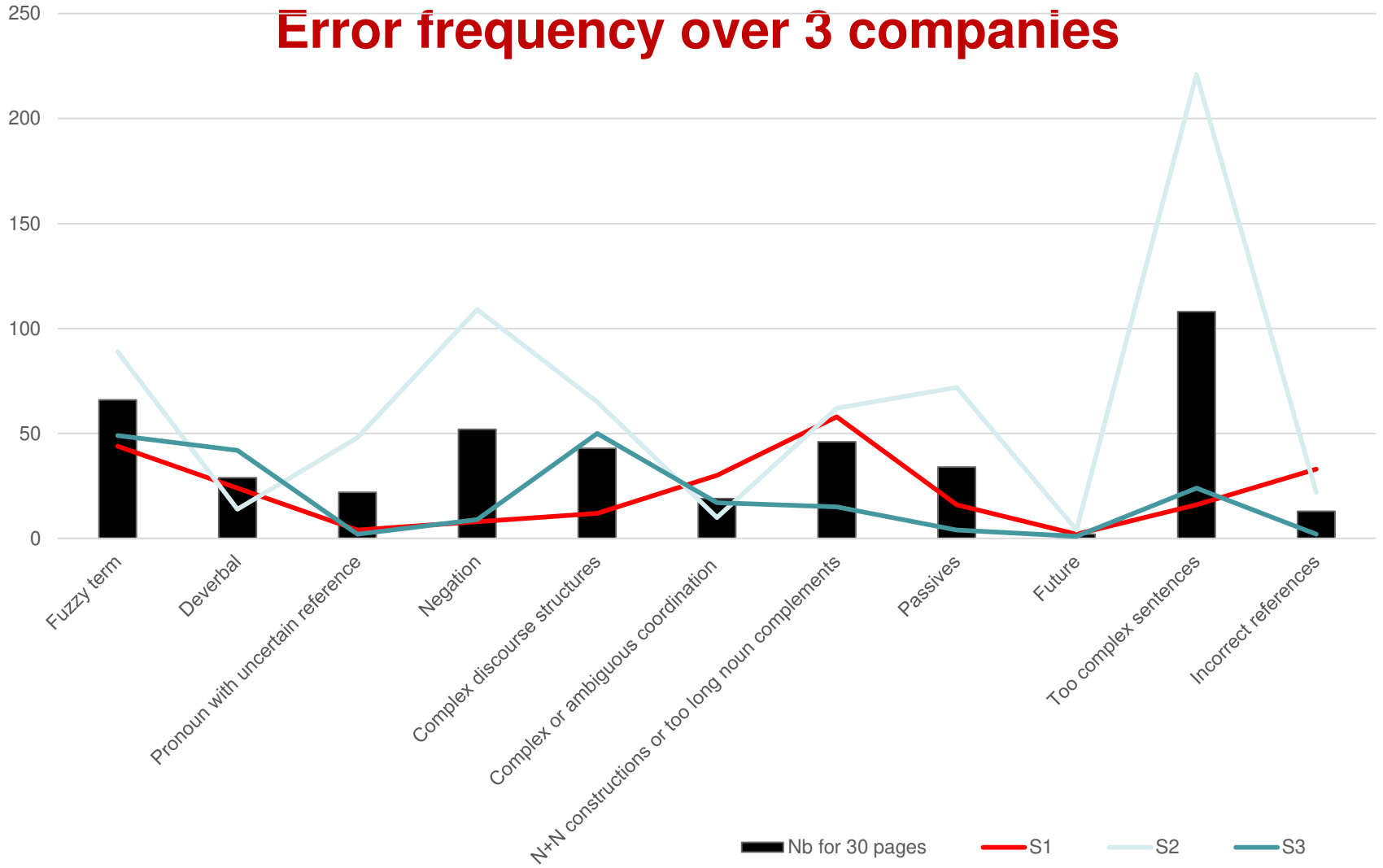
(2) Constructions complexes

if you encounter problems **when** working with TR eikon or receive error messages on **which** **CHECK: too many subordinates** diversifies you would like to obtain more information, you can contact support .

click market data **and** tools to view key data **and** access tools such as calculators **and** models **ADVICE: too many coordinations**

see introduction **to** system test running **from** the product **to** run system test **prior to** **CHECK too many subordinates** diversifies installation in the TR eikon system test guide.

Error frequency over 3 companies



Réactions des auteurs aux messages d'alerte

- (1) Fait une correction appropriée, la note pour la réutiliser
- (2) Approuve l'alerte mais découvre que l'erreur est plus large
- (3) Fait la correction, mais fait en même temps une autre erreur
- (4) Comprend l'erreur mais ne sait pas comment la corriger: ne fait rien pour l'instant ou demande de l'aide
- (5) Recherche des erreurs similaires pour voir comment elles ont été corrigées
- (6) Considère que l'erreur est mineure et ne la corrige pas.
- (7) Ne fait rien: erreur de diagnostic de LELIE: peut survenir par ex. Pour termes flous, negation, déverbaux si pas bien adapté au domaine.

Travaux en cours: aide à la correction des erreurs: quelques cas typiques

1. Modaux:

(a) Texte initial:

*Cependant, si aucune de ces solutions ne fonctionne, vous **pouvez** **ADVICE avoid modal** toujours contacter notre hotline....*

(b) Une correction est proposée par LELIE sur un fragment de texte plus large, sans perte de sens:

*Cependant, si aucune de ces solutions ne fonctionne,
[vous **pouvez** **ADVICE avoid modal** toujours contacter] → [**contactez**] notre hotline....*

(c) Après confirmation du rédacteur:

*Cependant, si aucune de ces solutions ne fonctionne, **contactez** notre hotline....*

■ Adverbes temporels flous:

La pression en vapeur doit être contrôlée régulièrement **ERROR: fuzzy term**

(a) Structure de correction suggérée:

La pression en vapeur doit être contrôlée

[régulièrement **ERROR: fuzzy term**] → [chaque/ toutes les T Unité]

(b) L'auteur peut alors compléter si OK:

- *La pression en vapeur doit être contrôlée [toutes les 20 minutes]*

3. Négation + événements complexes:

- *Ne jamais **ADVICE avoid negation** débrancher la machine **avant** que l'ADSL se soit arrêtée. **AVOID: two events in a requirement***

(a) Template de correction générique:

→ [Evenement_2 (présent)] puis [Evenement_1 (sans négation)]

(b) Correction de l'auteur:

- *Arrêter l'ADSL, puis débrancher la machine.*

→ Environ 60 templates de ce type définis pour diverses catégories d'erreurs.

Error type	Error pattern	Correction pattern	Example
Fuzzy determiner	[a few Noun]	[less than X Noun] *Adds an upper boundary X	A <u>few</u> minutes --> Less than 5 minutes
	[most Noun]	[more than X Noun] *Adds a lower boundary X	<u>Most</u> pipes shall ... →More than 8 pipes shall...
Temporal, iterative adverbs	[VP(action) Adverb(iterative)] *VP(action): action verb	[VP(action) every X(time)]	The steam pressure shall be controlled <u>regularly</u> →The steam pressure shall be controlled every 10 minutes.
Fuzzy prepositions	[near Noun(location)]	[less than X(distance) from Noun(location)] *X(distance) depends on Context	<u>Near</u> the gate → Less than 100 m from the gate
Negation on usages	[(do) not Verb(use) NP] *NP: any noun *Verb(use) any verb such as 'use'	[Verb(use) hyperonym(NP) other than NP] *Hyperonym(NP) denotes a more generic term than the NP, given in a domain terminology	<u>shall not use</u> hydrogen →shall use a gas other than hydrogen
Reverse synchronization	[do not/never VP before VP'] *VP and VP' denote two actions	[VP only after VP'] or [VP'. Then VP] *Actions are reversed in the correction, some persuasion effects may be lost.	<u>never unplug</u> before the machine has been stopped. →stop the machine. then unplug it

Sommaire de la présentation:

- Problèmes et défis en rédaction d'exigences en langue naturelle
- TextCoop et LELIE: 2 projet de recherche ANR
- La plateforme TextCoop pour l'analyse du discours technique
- LELIE: votre assistant pour écrire des exigences selon les normes

→ Tâches en cours et Projets

Emplois de TextCoop + LELIE

- Plug-In pour un système plus large de rédaction d'exigences, peut être associé à un éditeur de texte ou une suite éditoriale.
- Modulaire: les modules peuvent être utilisés indépendamment
- E/S en XML , adaptable à Word, Excel, pdf, Doors, etc.
- Adaptable à divers domaines et types de docs techniques
- TextCoop (env. 25 utilisateurs), LELIE (dans env. 7 sociétés)

→ **TextCoop V5.1 + LELIE V2.1: Open source.**

(contact stdizier@irit.fr)

→ Utilisable tel quel, ou comme des specs à recoder.

Travaux en cours et Perspectives

Fondements:

- Développement de l'analyse de la **cohésion**
- Développement d'aides spécifiques pour les francophones écrivant en anglais (projet IREB)
- Détection de formes d'**incohérence entre ensembles d'exigences**

Adaptation au domaine et à l'utilisateur:

- **Auto-adaptation** aux ressources métier (lexiques, savoir-faire)
- **Mémoire de correction** se souvenir des corrections réalisées en contexte et les proposer au rédacteur.

Analyse globale d'un texte (exigence de sécurité)

<main-requirement>

<title> **Monitoring safe operation of industrial trucks** </title>

<warning> Working practices must be monitored by a responsible supervisor
 <justification> to ensure that safe systems of work are followed. </justification>
</warning>

<purpose> This list is a basic guide for novices. </purpose>

<subtitle1> **Operators must always:** </subtitle1>

<requirement> observe floor loading limits - <restatement> in other words find
 out the weight of the laden truck. </restatement> </requirement>

<requirement> watch out before proceeding <justification> because of
 pedestrians and bystanders <illustration>(see paragraph p. 390-394)
 </illustration>. </justification> </requirement>

<warning> travel on soft ground <concession> unless the industrial truck is
 suitable for this purpose </concession>. </warning>

<warning> never allow unauthorised people to operate the truck. </warning>

..... </main-requirement>

Exemple de règle et évaluation

Contrast → conn(opposition whe), gap(G), ponct(comma). /

conn(opposition whe), gap(G), end. /

conn(opposition how), gap(G), end.

conn(opposition whe): *whereas, but whereas, but while*

conn(opposition how): *however.*

structure	number manually annotated	precision (%)	recall (%)
justification	49	92	86
reformulation	27	85	76
illustration	112	91	82
precision	52	90	85
comparison	58	89	85
consequence	12	92	87
contrast	49	86	79
concession	35	89	82

Table 1. Textcoop performances for the identification of discourse relations