

Passer du langage au modèle

Sujet Procédés de modélisation sémantique

Objet du procédé Exploiter un texte pour alimenter le modèle

Mots clefs aspect sémantique, grammaire, linguistique, langage naturel Praxeme, méthode, procédé

Référence **PxPCD-21**

État à valider

Version 0.1.0

Date 1er mai 2013

Auteurs, contributeurs Dominique VAUQUIER

Relecteurs

Sommaire

1. CONTEXTE D'APPLICATION DU PROCÉDÉ.....	3
1.1 Objet.....	3
1.2 Situations d'usage	3
1.3 Positionnement dans la méthode	3
1.4 Conditions à respecter	3
2. TERMINOLOGIE EMPLOYÉE	3
2.1 Champ sémantique	3
2.2 Réseau sémantique	4
2.3 Analyse textuelle.....	4
2.4 Analyse componentielle	4
2.5 Correspondance syntaxique.....	4
2.6 Terminologie linguistique et logique	4
3. COMPÉTENCES REQUISES.....	4
4. MODE OPÉRATOIRE.....	4
4.1 Dessiner les réseaux sémantiques.....	4
4.2 Mener une analyse textuelle	5
4.3 Reformuler les propositions	7
5. RÉSULTATS PRODUITS.....	7
5.1 Éléments candidats.....	7
5.2 Liens de traçabilité	7
6. OUTILLAGE DU PROCÉDÉ.....	7
7. APPROFONDISSEMENTS.....	7
7.1 Correspondances avec d'autres référentiels	7
7.2 Bibliographie pratique.....	7

Rappels méthodologiques

Dans le contexte de la méthode Praxeme, un *procédé* est « une façon de faire, un mode opératoire pour exécuter une tâche »¹. Il s'agit donc d'une prescription à un niveau individuel, par opposition au *processus* qui est une réponse méthodologique au niveau collectif.

Les fiches de procédés ne font pas référence à d'éventuels processus dans lesquels ces procédés pourraient intervenir, ceci afin de faciliter leur réemploi dans plusieurs contextes.

Protection du document

L'initiative pour une méthode publique repose sur le bénévolat et la mutualisation des investissements entre ses contributeurs. Elle vise à élaborer et à diffuser une méthode ouverte et libre de droits. Sa dynamique n'est possible que si cet esprit est maintenu à travers les utilisations des documents qu'elle met à la disposition du public. C'est pourquoi les documents sont protégés par une licence « *creative commons* »² qui autorise l'usage et la réutilisation de tout ou partie d'un document du fonds Praxeme sous seule condition que l'origine en soit citée. Les éventuels documents dérivés, qui reprennent du contenu de Praxeme, doivent s'appliquer à eux-mêmes les mêmes conditions, faire référence à la « *creative commons* » et porter les symboles idoines :



Actualisation de ce document

Pour obtenir la dernière version de ce document, se rendre sur le wiki du *Praxeme Institute*, à la page : <http://www.praxeme.org/index.php?n=Modus.PxPCD21LangMdl>.

L'historique du document

Indice	Date	Rédacteur	Contenu
0.0.0	01/05/2013	DVAU	Première rédaction
0.1.0	01/05/2013		Version actuelle du document

¹ Cf. rubrique Thesaurus sur le site du *Praxeme Institute* : <http://www.praxeme.org/index.php?n=Thesaurus.Procedure>.

² Voir la philosophie et le détail des licences sur : <http://creativecommons.org/>.



Pour suivre l'actualité de la méthode publique

- Mailing list
- Groupe LinkedIn
- Twitter
- le wiki

Pour participer aux travaux du Praxeme Institute

- Adhésion au *Praxeme Institute*

<http://www.praxeme.org/index.php?n=Chorus.Join>

1. Contexte d'application du procédé

1.1 Objet

Ce procédé « Passer du langage au modèle » a pour but de :

« **Exploiter un texte pour alimenter le modèle.** »

1.2 Situations d'usage

Tout texte, de la phrase au document, de l'exigence élémentaire au manuel de procédures, peut se prêter à cet exercice.

L'application stricte, pas à pas, du procédé se justifie pour les points sensibles dans l'analyse du métier ou lors de controverses.

Une mise en œuvre plus lâche peut s'avérer utile pour démarrer une modélisation dans un domaine que l'on ne connaît pas.

Une autre situation est celle de l'atelier avec des représentants du domaine. Ils formulent une assertion de la connaissance métier. L'animateur la transcrit scrupuleusement au tableau. Puis le modélisateur applique le procédé devant le groupe référent. De cette façon, on n'arrive pas forcément au modèle définitif, mais cela sert à soulever des questions, à révéler des contradictions. Souvent, cet exercice oblige à creuser plus profond.

1.3 Positionnement dans la méthode

a. Place dans le cadre de référence

Les textes en entrées peuvent provenir de n'importe quel aspect du Système Entreprise. La source la plus naturelle est l'aspect intentionnel qui agglomère les expressions qui définissent et détaillent la mission de l'entreprise.

Le procédé lui-même est de nature linguistique. Il produit des éléments de modélisation qui peuvent être dirigés vers n'importe quel aspect, selon la nature du texte.

b. Relations avec d'autres procédés

Ce procédé est particulièrement lié aux procédés de modélisation sémantique de portée « classe »³.

c. Posture

Il s'agit d'un procédé d'analyse, « collé » sur le matériau d'entrée.

1.4 Conditions à respecter

Nous ne nous intéressons ici qu'aux éléments destinés à l'aspect sémantique. Le modélisateur doit donc éliminer toutes les notions qui n'entreraient pas dans la sphère des fondamentaux du métier. Par exemple, beaucoup de termes peuvent renvoyer à des acteurs ou des objets de nature organisationnelle. L'analyse textuelle peut les conserver mais leur destination est l'aspect pragmatique.

Par ailleurs, il faut garder présent à l'esprit que l'application servile de ce procédé ne peut suffire pour construire un « bon » modèle sémantique. Il convient de n'y voir qu'une aide pour démarrer la modélisation.

2. Terminologie employée

Les paragraphes ci-dessous donnent les définitions des termes techniques utilisés dans cette fiche. La fiche PxPCD-20 fournit la terminologie plus générale, pour l'ensemble des procédés sémantiques.

2.1 Champ sémantique

Ensemble des différents sens d'un mot.

³ Les procédés de PxPCD-22 à PxPCD-24. Voir l'organisation des procédés sémantiques dans la fiche chapeau PxPCD-20.

2.2 Réseau sémantique

Représentation, souvent graphique, des relations entre des concepts.

2.3 Analyse textuelle

Examen minutieux d'un texte, mettant en jeu une ou plusieurs approches linguistiques.

2.4 Analyse componentielle

Décomposition d'un signifié en ses sèmes élémentaires.

2.5 Correspondance syntaxique

Correspondance entre catégories de deux systèmes de communication obéissant chacun à une syntaxe.

Une syntaxe est un ensemble de règles formelles qui permettent la construction d'énoncés dans un système de communication. La langue est un tel système ; de même, UML.

2.6 Terminologie linguistique et logique

a. Déictique

Mot qui renvoie à une autre expression, normalement antérieure dans le texte.

Exemples : « ceci », « celui-ci », les pronoms, certains adverbes...

b. Embrayeur

Terme qui fait le lien entre l'énoncé (le texte) et l'énonciation (la situation dans laquelle le texte a été produit).

On les rencontre dans les comptes rendus ou dans les textes réglementaires. Le modélisateur doit y prêter attention car, d'une part, ces embrayeurs créent un champ qui polarise la signification et, d'autre part, les circonstances de l'énonciation accordent le statut du texte, son degré de vérité (est-il plus ou moins incontournable, contraignant ?).

c. Foncteur

Catégorie (ici de terme ou d'expression) qui assure une fonction dans le phénomène étudié (ici la signification).

d. Joncteur

Catégorie (de terme ou d'expression) qui permet de relier plusieurs notions autonomes.

3. Compétences requises

Pour mettre en œuvre ce procédé, il suffit d'avoir un peu de bon sens, une connaissance minimale de la grammaire ainsi que la familiarité avec les catégories de représentation de l'aspect sémantique. Ces dernières sont exposées dans le guide « Approche de l'aspect sémantique ».

4. Mode opératoire

4.1 Dessiner les réseaux sémantiques

Avant d'être en mesure d'élaborer un premier modèle, l'analyste peut adopter une approche moins formelle. La technique des champs sémantiques lui permet de recueillir et d'organiser, en première approche, la terminologie du domaine étudié. Cette technique, très simple, est moins formelle que la modélisation. On peut la voir comme une étape intermédiaire. Elle consiste à jeter, sur le papier, les termes qui apparaissent dans l'univers du discours et à les relier en fonction de leur proximité sémantique.

Dans cette première passe, on ne dispose donc que de deux catégories de représentation :

- les termes ;
- les relations entre ces termes (des flèches éventuellement assorties d'un libellé qui exprime la nature de la relation entre les deux termes).

Le réseau sémantique tolère les synonymes. C'est plus une technique terminologique, proche du thesaurus, et qui pose une pile de pont entre le matériau brut et le premier modèle. Elle permet une première organisation de la matière que le modélisateur reprend ensuite. Pour passer des réseaux sémantiques au modèle, le modélisateur doit statuer sur chaque terme, éliminer les synonymes et la redondance, déterminer le type d'élément de modélisation qui convient le mieux.

4.2 Mener une analyse textuelle

a. Principe

La technique suivante est celle de l'analyse textuelle. Elle repose sur l'hypothèse de la correspondance syntaxique : correspondance entre les catégories du langage naturel et celles retenues par l'approche de l'aspect sémantique. L'analyse textuelle permet de passer, presque mécaniquement, des textes recueillis vers le modèle, du moins elles proposent des éléments candidats pour le modèle.

Cette technique fournit un bon point de départ pour un modélisateur qui se sent un peu perdu face à la complexité ou à la nouveauté du domaine étudié. Elle incite à collecter ou à rédiger les descriptions du domaine.

b. Correspondance syntaxique

UML est une syntaxe un peu plus formelle que le langage naturel (pas complètement formelle, cependant). La correspondance entre les catégories syntaxiques des deux langages offre un guide pratique pour le modélisateur, ainsi qu'une preuve de la fidélité du modèle. C'est ce que tente le tableau ci-dessous.

Ce tableau de correspondance syntaxique est un premier essai. Il faudrait sans doute aller plus loin et s'appuyer sur ce que nous apprend la linguistique.

La première colonne représente un essai de classification en termes de fonctions logiques du discours.

Figure PxPCD-21_1. Le tableau de correspondance syntaxique

Catégorie ou foncteur (essai)	Catégorie du langage naturel	Élément de modélisation	Recommandation
« Référenceur » en extension et intension	Nom, groupe nominal en position de sujet ou d'objet	Classe	Se méfier des formes substantivées. Elles cachent l'action. Mieux vaut retenir, dans ce cas, le verbe.
« Joncteur »	Nom, groupe nominal en position de complément, préposition	Rôle sur une association	La mise en relation entre deux objets doit être restituée par une association, plutôt que par un nom d'attribut.
Information	Nom caractérisant des valeurs variables	Attribut (éventuellement calculé ou de portée classe), opération quand il y a besoin de paramètres	Éviter les attributs de type Classe (à l'exception des classes utilitaires), leur préférer des associations qui disent davantage sur la connexion entre les concepts.
Qualificatif	Adjectif qualificatif (associé à un substantif repris par une classe), auxiliaire « être » suivi d'un participe passé ou d'un adjectif	État dans une machine à états (automate à états)	Les objets « métier » centraux ont souvent un cycle de vie. Un modèle sans automate est suspect.
Action	Verbe d'action dont le sujet ou l'objet est un objet candidat ⁴	Opération ou association si l'action implique deux objets candidats ou plus	La question est « où placer l'opération ? » : sur l'objet qui agit (sujet) ou sur celui qui subit (complément d'objet direct). ⁵
« Joncteur » expression de structure	Verbe d'état (constituer, se composer, contenir...)	Association, souvent assemblage ou composition ; parfois, état	La composition ou l'assemblage n'interdit pas de nommer l'association.
Action et qualificatif	Auxiliaire « être » suivi d'un participe présent	État avec une action	Mot réservé « do » dans le dessin de l'état.
« Joncteur » ou qualificatif	Auxiliaire « avoir »	Attribut ou assemblage ou association s'il existe un terme plus expressif	Selon le complément d'objet.
Classificateur	« est un », « est une sorte de », « spécialise »...	Héritage	Soulève plusieurs problèmes : classification multicritères, mutation...
Coordinateur	Conjonction de coordination	Synchronisation dans un processus ou un automate	« et », « ou », « donc ». Ne pas réifier les processus.
Embrayeur, déictique			Toujours vérifier que la référence est claire. Éventuellement, reformuler.

⁴ Cela évacue les actions dont le sujet est un acteur et qui sont donc de nature pragmatique.

⁵ Le critère pour trancher cette question est l'appartenance sémantique. De quel concept l'opération est-elle caractéristique ?

4.3 Reformuler les propositions

Pour que l'analyse textuelle se fasse au mieux, il est parfois nécessaire d'intervenir sur les textes, soit pour les restructurer et les rendre plus faciles à manipuler, soit pour les désambiguïser. La désambiguïstation porte souvent sur les déictiques, dont on perd le référent dans des textes trop chargés.

La rédaction de ces descriptions préalables obéit à des règles de style : phrases simples (sujet, verbe, complément) ; modalités clairement exprimées ; articulations sans ambiguïté... Quelques conseils :

- Passer de la forme passive à la forme active.
- Faire des phrases plus courtes.
- Expliciter les objets singuliers, les déictiques et les sous-entendus...

5. Résultats produits

5.1 Éléments candidats

Ce procédé produit des listes de termes candidats pour le modèle, ainsi que leurs relations.

Une façon de faire est de les enregistrer directement dans le modèle :

- soit dans un paquetage « bac à sable » où ils pourront être recherchés et d'où on pourra les déplacer vers le bon endroit du modèle, si on décide de les conserver ;
- soit directement dans le modèle mais en les marquant comme « candidat », tant que la décision de les conserver n'a pas été entérinée.

Ce marquage se fait facilement au moyen d'une annotation au sens d'UML (*tagged value*).

5.2 Liens de traçabilité

La qualité attendue d'un modèle sémantique est sa fidélité à l'univers du discours. Elle se prouve par la capacité du modèle à reprendre le maximum des unités de sens présentes dans cet univers.

Par ce procédé, le modélisateur met en place les chaînes de traçabilité qui « justifie » les éléments de modélisation, c'est-à-dire qui réfère les éléments aux endroits précis dans les matériaux d'entrée où il a trouvé leur manifestation.

6. Outillage du procédé

Plusieurs outils reposent sur ce principe.

Manuellement, le modélisateur peut procéder en surlignant avec des couleurs différentes les éléments de texte en fonction de leur catégorie : substantifs (candidats pour une classe), verbes de relation (associations), verbes d'action (opérations), adjectifs et participes (états), etc.

7. Approfondissements

7.1 Correspondances avec d'autres référentiels

Néant.

7.2 Bibliographie pratique

Le Génie cognitif, Claude Vogel

Le Langage, Bibliothèque de la Pléiade.

Psychologie, Bibliothèque de la Pléiade. Voir particulièrement le chapitre : Les fonctions de stockage, p. 1836.

Approche de la linguistique, Véronique Schott-Bourget.