

Lexical Resource Semantics

Grégoire Winterstein

Avril 2008

Survol des caractéristiques de *LRS* :

- Formalisme d’interface syntaxe-sémantique : traitement de la compositionnalité
- Pas de langage intermédiaire avec la sémantique : les objets manipulés sont des objets sémantiques (formules logiques, DRS, ...)
- Cohérent avec les mécanismes HPSG : usage crucial de la sous-spécification et de l’unification

1 Motivations de LRS

Arguments présentés dans : [Richter and Sailer, 2003].

- Des systèmes comme *MRS* n’arrivent pas à exclure des phrases “agrammaticales” pour des raisons de conflits de scope de quantifieurs (cf. section 1.1)
- Des systèmes comme *Flexible Montague Grammar* ont une bonne couverture empirique mais sont complexes et peu naturels à implémenter en HPSG (cf. section 1.3)

1.1 Le subtil argument de la négation polonaise (et anglaise)

Avec une sémantique du type de *MRS*, si l’on désire exclure certains énoncés dans la grammaire :

- L’ambiguïté de la marque de la négation en polonais ne peut être exprimée par une sous-spécification de la contribution sémantique
- L’ambiguïté de la marque de la négation en anglais ne peut être exprimée par une ambiguïté lexicale

1.1.1 Données

En polonais (cf.[Przepiórkowski and Kupść, 1999]), *nie*, la marque de la négation, est ambiguë :

- négation d’éventualité : (1)
- négation de non-éventualité (pléonastique) : (2)

(1) *Jej nie przewróciłem*
her NM I overturned
Je ne l’ai pas renversée

(2) *Omal jej nie przewróciłem*
almost her NM I overturned
Je l’ai presque renversée

En anglais, la négation est également ambiguë :

- négation d'éventualité : (3)
- négation métalinguistique (obligatoire pour interpréter les énoncés avec PPI comme (4))

(3) Peter didn't leave

(4) (*) Peter didn't already leave

1.1.2 Argumentation

Il est impossible d'avoir un traitement unifié pour la marque de la négation (sous-spécification ou ambiguïté lexicale).

Les énoncés inacceptables seront bien exclus, mais au niveau de la récupération de la forme logique, pas au niveau de la grammaire.

Anglais *Hypothèse A* : la marque de négation est lexicalement ambiguë, (4) reçoit deux analyses, une par entrée lexicale. Idéalement la grammaire devrait exclure l'analyse avec la négation d'éventualité.

- En (4) la négation d'éventualité est exclue par un conflit de portée :
 - *already* doit prendre portée hors de la négation (cf. PPI)
 - *already* doit modifier une éventualité résultative : impossible avec une portée sur la négation
- Comme les portées respectives des items ne sont pas résolues dans la grammaire on ne peut y exclure l'analyse de (4) avec négation d'éventualité

→ On associe une seule entrée sous-spécifiée à la marque de négation. La procédure de résolution sémantique fournira l'unique bonne lecture.

Polonais *Hypothèse B* : la marque de la négation possède une seule entrée lexicale avec un contenu sémantique sous-spécifié.

- En polonais : les *n-words* requièrent la présence d'un *nie* d'éventualité (*neg-concord*, cf. section 3.2)
- En (5) il y a un conflit entre les contraintes relatives à *Omal* (\Rightarrow *nie* pléonastique) et à *nikogo* (\Rightarrow *nie* d'éventualité)
- Comme la représentation de *nie* est sous-spécifiée on ne peut exclure (5) dans la grammaire (mais la résolution de la forme logique sera impossible)

(5) ?* *Omal nikogo nie przewróciłem*
almost nobody NM I overturned

J'ai presque renversé personne (?)

Conclusion Avec une interface à la *MRS* aucune solution générale n'est satisfaisante pour un traitement général de la marque de négation.

→ Changeons de sémantique.

1.1.3 Solution 1 : expliciter les portées

Puisque ce sont les conflits de portée qui excluent certains énoncés on peut choisir d'indiquer les relations de portée ("pluggings") dans la grammaire.

- On abandonne le bénéfice des dénotations sous-spécifiées, les structures diffèrent par leur trait *CONT*
- On conserve les difficultés techniques de l'approche (cf. section 1.2)

1.2 Difficultés techniques supplémentaires

- La complexité à donner à l’expression des contraintes de portée est incertaine (on peut imaginer des contraintes de la forme *si l_1 est une sous-expression de l_2 alors l_3 est une sous-expression de l_4* , par exemple pour le cas des portées relatives des pronoms interrogatifs en allemand)
- La bonne formation des jeux de contraintes (non-cyclicité etc.) est difficile à exprimer, même si techniquement RSRL le permet (impossible en *MRS* seul)
- Le domaine empirique de la sous-spécification n’est pas proprement déterminé :
 - rattachement des *PP* : sous-spécification du type des arguments ? (mais alors une seule analyse syntaxique)
 - lecture distributive/collective de certains quantifieurs : jusqu’ici jamais sous-spécifié

1.3 Solution 2 : λ -calcul et type-shift

LF – Ty2 : adaptation de la *Flexible Montague Grammar* de [Hendriks, 1993], la solution classique pour traiter les ambiguïtés sémantiques, développée dans [Sailer, 2000].

- Abandon total de toutes les formes de sous-spécification
- Utilisation du λ -calcul, compositionnalité = application fonctionnelle
- Un mot possède une entrée lexicale unique et continue, de laquelle on peut en dériver d’autres par “type-shifting”

Avantages

- pas de problèmes techniques comme en section 1.2
- pas de problèmes empiriques (mais voir en section 3.2)

Désavantages

- beaucoup de dérivations spurieuses pour une lecture similaire (du fait de la grande variété de type-shift possibles)
- les processus de type-shifting sont complexes et peu naturels en *HPSG*

1.4 Comparaisons des caractéristiques

Variation de 3 paramètres

1. *Dénotation sous-spécifiée* : la valeur sémantique d’un syntagme est en partie indéterminée, la grammaire ne fournit pas directement les différentes lectures qui sont calculées en aval
2. *Représentation indirecte* : occurrences de méta-variables, utilisation d’un langage intermédiaire entre syntaxe et sémantique
3. *Représentation discontinue* : la forme logique associée à un signe n’est pas une formule continue mais une collection d’expressions entretenant d’éventuelles relations de hiérarchie

Approche	dénotation sous-spécifiée	représentation indirecte	représentation discontinue
<i>Ty2U / MRS</i>	+	+	+
<i>Ty2U^P</i>	-	+	+
<i>LF – Ty2</i>	-	-	-
<i>LRS</i>	-	-	+

Aux avantages de *LF – Ty2* on cherche à ajouter ceux de *Ty2/MRS* : *LRS* utilise une représentation discontinue.

2 Les grandes lignes de *LRS*

Les objets sémantiques manipulés par *LRS* sont des formules d'un langage sémantique donné, ici *Ty2* (proche de la logique montagovienne)

- Une structure entièrement résolue correspond à une lecture donnée
- La compositionnalité fournit des formules (potentiellement) sous-spécifiées (*i.e.* ayant des types non maximalelement spécifiés)
- Les contraintes sur la bonne formation des formules sont exprimées en *RSRL*, un langage formel de description de *HPSG*

2.1 Structure

$$\left[\begin{array}{l} lrs \\ EXCONT \ me \\ INCONT \ me \\ PARTS \ list(me) \end{array} \right]$$

- *EXCONT* et *INCONT* correspondent (grossièrement) aux attributs *GTOP* et *LTOP* de *MRS*
 - *LTOP* correspond (généralement) à la principale contribution sémantique de l'élément
- *PARTS* : liste des contributions sémantiques d'un mot ainsi que de toutes leurs sous-expressions déjà partiellement spécifiées

2.2 Combinatoire

2.2.1 Bonne formation des *lrs*

Deux principes :

(6) Le principe *INCONT* :

1. Dans chaque *lrs* la valeur de *INCONT* est un élément de *PARTS* et une sous-expression de *EXCONT*

(7) Le principe *EXCONT* :

1. dans chaque syntagme la valeur du trait *EXCONT* de la non-tête est un élément du trait *PARTS* de la non-tête
2. dans chaque phrase toute sous-expression de la valeur de *EXCONT* est un élément du trait *PARTS* de la phrase, et tout élément de *PARTS* est une sous-expression de *EXCONT*

2.2.2 Compositionnalité

(8) Dans chaque syntagme avec tête :

1. les traits *EXCONT* du syntagme et de sa tête sont identiques
2. les traits *INCONT* du syntagme et de sa tête sont identiques
3. le trait *PARTS* du syntagme contient tous, et uniquement, les éléments des traits *PARTS* de ses fils
4. les conditions suivantes doivent être respectées :
 - si la non-tête est un quantifieur alors

- son trait *EXCONT* prend la valeur $\left[\begin{array}{l} \textit{quantifier} \\ \text{VAR } \textit{var} \\ \text{SCOPE } \left[\begin{array}{l} \textit{l-const} \\ \text{ARG1 } \rho \\ \text{ARG2 } \nu \end{array} \right] \end{array} \right]$
- la valeur α du trait *INCONT* de la tête du syntagme est un composant de ρ : $\alpha \triangleleft \rho$ (=restriction)
- les valeurs *EXCONT* des deux fils sont identiques (permet le liage des variables)
- si la non-tête est un NP quantifié avec une trait *EXCONT* de la forme $\left[\begin{array}{l} \textit{quantifier} \\ \text{VAR } \textit{var} \\ \text{SCOPE } \left[\begin{array}{l} \textit{l-const} \\ \text{ARG1 } \rho \\ \text{ARG2 } \nu \end{array} \right] \end{array} \right]$ alors la valeur α du trait *INCONT* de la tête est une sous-expression de ν : $\alpha \triangleleft \nu$ (=portée)

2.2.3 Contraintes sur la forme logique

Les contraintes du principe sémantique sont exprimées en *RSRL* :

- *RSRL* est un langage de description de *HPSG*
- la notion de *sous-composant* (\triangleleft) est formellement définie en *RSRL*
- les contraintes du type $\alpha \triangleleft \beta$ ne font pas partie du signe
- ce sont des contraintes sur l'objet de l'analyse, au même titre que le signe contraint cet objet en spécifiant le type de ses traits

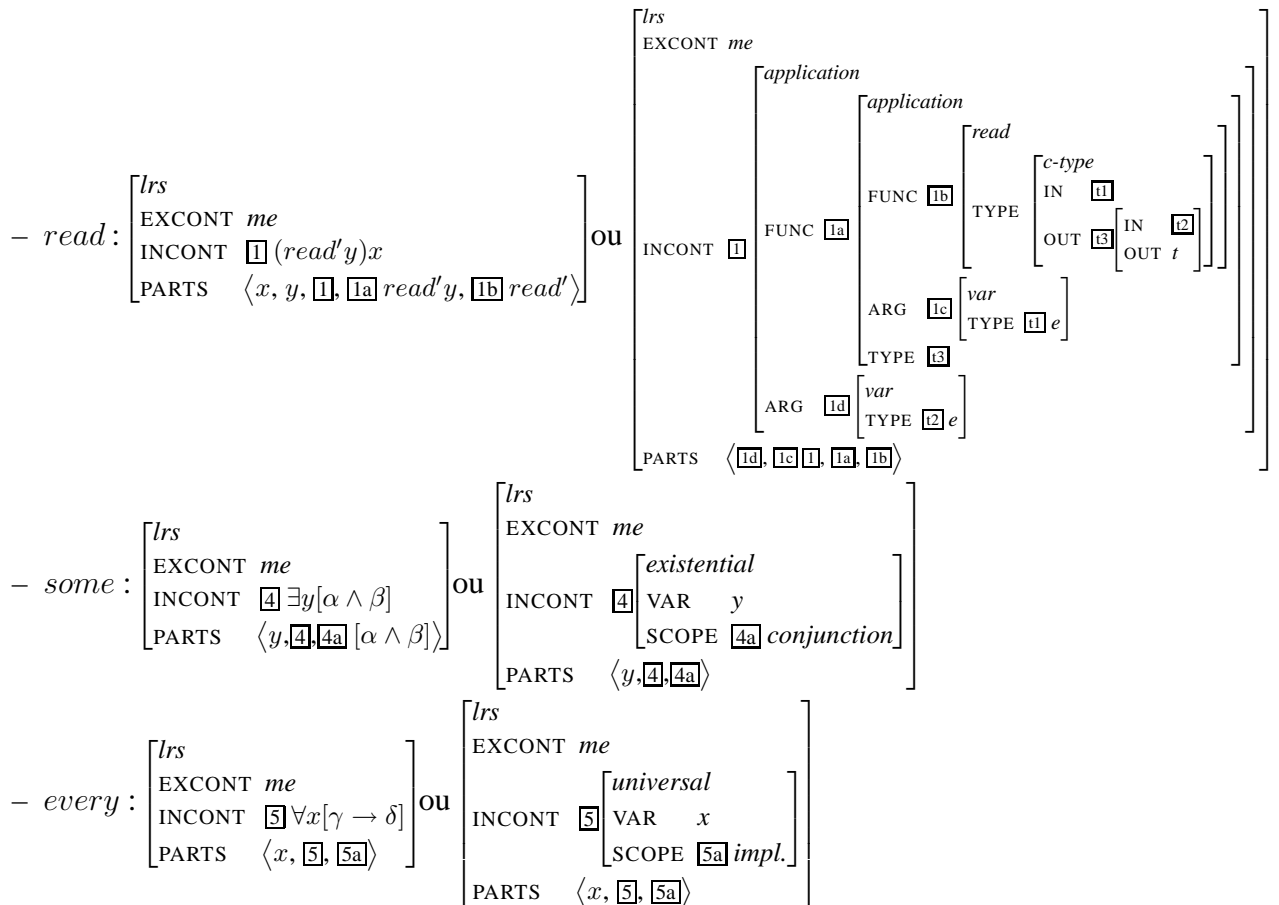
3 Applications

3.1 Ambiguïté de portée

(9) Every student reads some book

3.1.1 Entrées lexicales

- *student* : $\left[\begin{array}{l} \textit{lrs} \\ \text{EXCONT } \left[\begin{array}{l} \textit{quantifier} \\ \text{VAR } x \end{array} \right] \\ \text{INCONT } \boxed{3} \textit{student}'x \\ \text{PARTS } \langle x, \boxed{3}, \boxed{3a} \textit{student}' \rangle \end{array} \right]$ ou $\left[\begin{array}{l} \textit{lrs} \\ \text{EXCONT } \left[\begin{array}{l} \textit{quantifier} \\ \text{VAR } x \end{array} \right] \\ \text{INCONT } \boxed{3} \left[\begin{array}{l} \textit{application} \\ \text{FUNC } \boxed{3a} \left[\begin{array}{l} \textit{student} \\ \text{TYPE } \left[\begin{array}{l} \textit{c-type} \\ \text{IN } \boxed{e} \\ \text{OUT } \textit{truth} \end{array} \right] \end{array} \right] \\ \text{ARG } \boxed{3b} \left[\begin{array}{l} \textit{var} \\ \text{TYPE } \boxed{e} \textit{entity} \end{array} \right] \end{array} \right] \\ \text{PARTS } \langle \boxed{3b}, \boxed{3}, \boxed{3a} \rangle \end{array} \right]$
- *book* : $\left[\begin{array}{l} \textit{lrs} \\ \text{EXCONT } \left[\begin{array}{l} \textit{quantifier} \\ \text{VAR } y \end{array} \right] \\ \text{INCONT } \boxed{2} \textit{book}'y \\ \text{PARTS } \langle y, \boxed{2}, \boxed{2a} \textit{book}' \rangle \end{array} \right]$



3.1.2 Composition

Voir Figure 1 en page 9

3.1.3 Lectures

- $\boxed{6}$ s'identifie à $\boxed{5}$: $\forall \exists$
- $\boxed{6}$ s'identifie à $\boxed{4}$: $\exists \forall$

⇒ La mécanique reste proche de *MRS*

3.2 Polonais et Neg-Concord

Une utilisation plus subtile de *LRS*.

3.2.1 Données

- la négation de phrase est exprimée par *nie* : (10)

(10) *Janek nie pomaga ojcu*
 Janek.NOM NM aide père.DAT
 Janek n'aide pas son père

- la présence de *nie* est obligatoire avec les *n-word* : (11), (12) ...

(11) *Janek *(nie) pomaga nikomu*
 Janek.NOM (NM) aide personne.DAT
 Janek n'aide personne

- (12) *Nikt* **(nie) pomaga ojcu*
 nobody.NOM (NM) aide père.DAT
 Personne n'aide son père
- Analyse traditionnelle : les *n-words* n'ont pas de négation inhérente
 - Un argument parmi d'autres (cf. [Richter and Sailer, 2003]) pour une analyse avec négation : (14) vs. (15)
- (13) *Ile* *przeczytales książek?*
 Combien tu as lu livres
 Combien t'as lu de livres ?
- (14) *Żadnej.* / *Żadną*
 Aucun.GEN / Aucun.ACC
- (15) *Nie przeczytałem [żadnej książki] / *[żadną książkę]*
 NM j'ai lu [aucun livre].GEN / [aucun livre].ACC
 Je n'ai lu aucun livre

Question

- si *nie* et les *n-words* apportent chacun une négation pourquoi n'y a-t-il pas de double négation en (11) ?

3.2.2 Analyse

LRS permet d'unifier certaines contributions *n-words*. Voir Figure 2 en page 9.

Résultats L'analyse donne trois valeurs sémantiques possibles :

- (16) $\neg\neg\exists x[human'(x) \wedge come'(x)]$ ($\boxed{4} < \boxed{2} = \boxed{0}$) = $\exists x[human'(x) \wedge come'(x)]$
 (17) $\neg\exists x\neg[human'(x) \wedge come'(x)]$ ($\boxed{2} < \boxed{4} = \boxed{0}$) = $\forall x[human'(x) \wedge come'(x)]$
 (18) $\neg\exists x[human'(x) \wedge come'(x)]$ ($\boxed{4} = \boxed{2} = \boxed{0}$)

On ne veut que (18), on postule la contrainte suivante pour le polonais :

- (19) Contrainte de Complexité de la Négation (*NCC*) :
 Chaque signe peut avoir au plus une seule négation qui est un composant de son trait *EXCONT* et a comme composant le trait *INCONT*

Cette contrainte n'est pas totalement *ad-hoc* :

- La concordance négative est obligatoire en polonais : *NCC* apparaît justifiée
- D'autres contraintes similaires ont été proposées pour les langues romanes : [Corblin and Tovina, 2000]
- \Rightarrow "Grammaticalisation" du nombre de négations pouvant apparaître dans un énoncé
- *NCC* est équivalent à la *negation absorption* de [Haegeman and Zanuttini, 1991] (avec l'avantage de reposer sur des mécanismes naturels du formalisme)

Neg-concord

- Le principe *NCC* impose la lecture en Neg-concord
- Mais il n'impose pas la présence de *nie* avec les *n-word* : (20) est accepté, avec la structure en Figure 3 en page 9.

- (20) * *Nikt* *przyszedł*
 Personne est venu

On postule une deuxième contrainte, encore sur le modèle de [Haegeman and Zanuttini, 1991] et [Corblin and Tovenà, 2000] (mais sur une base sémantique) :

(21) Critère de Négation (*NegC*) :

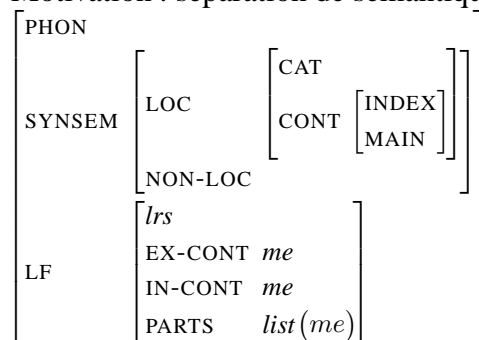
Pour chaque verbe, s'il y a une négation dans le trait *EXCONT* du verbe qui a portée sur le trait *INCONT* du verbe, alors la négation doit appartenir au trait *PARTS* du verbe

– (20) serait acceptable si le verbe *przyszedeł* contenait une négation (avec l'affixe *nie*)

4 *LRS* dans *HPSG*

Défendu dans [Sailer, 2004].

Motivation : séparation de sémantique lexicale et compositionnelle : *CONT* vs. *LF*.



– permet de conserver la sémantique de Pollard et Sag 94

– permet de pouvoir exprimer la contribution sémantique principale d'un item (trait *MAIN*)

– ...

Références

- [Corblin and Tovenà, 2000] Corblin, F. and Tovenà, L. (2000). On the multiple expression of negation in romance. In d'Hulst, Y., Rooryck, J., and Schrotten, J., editors, *Romance Languages and Linguistic Theory 1999*, pages 87–115. John Benjamin, Amsterdam.
- [Haegeman and Zanuttini, 1991] Haegeman, L. and Zanuttini, R. (1991). Negative heads and the neg criterion. *The linguistics review*, 8 :233–251.
- [Hendriks, 1993] Hendriks, H. (1993). *Studied Flexibility*. PhD thesis, Institute for Logic Language and Computation, Amsterdam.
- [Przepiórkowski and Kupść, 1999] Przepiórkowski, A. and Kupść, A. (1999). Eventuality negation and negative concord in polish and italian. In Borsley, R. and Przepiórkowski, A., editors, *Slavic in HPSG*, pages 211–246. CSLI Publications.
- [Richter and Sailer, 2003] Richter, F. and Sailer, M. (2003). Basic concepts of lexical resource semantics. Author material for a course on Semantics (ESSLLI 2003).
- [Sailer, 2000] Sailer, M. (2000). *Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar*. PhD thesis, Eberhard-Karls-Universität Tübingen.
- [Sailer, 2004] Sailer, M. (2004). Local semantics in head-driven phrase structure grammar. In Bonami, O. and Hofherr, P. C., editors, *Empirical issues in Syntax and Semantics*, volume 5, pages 197–214. CNRS.

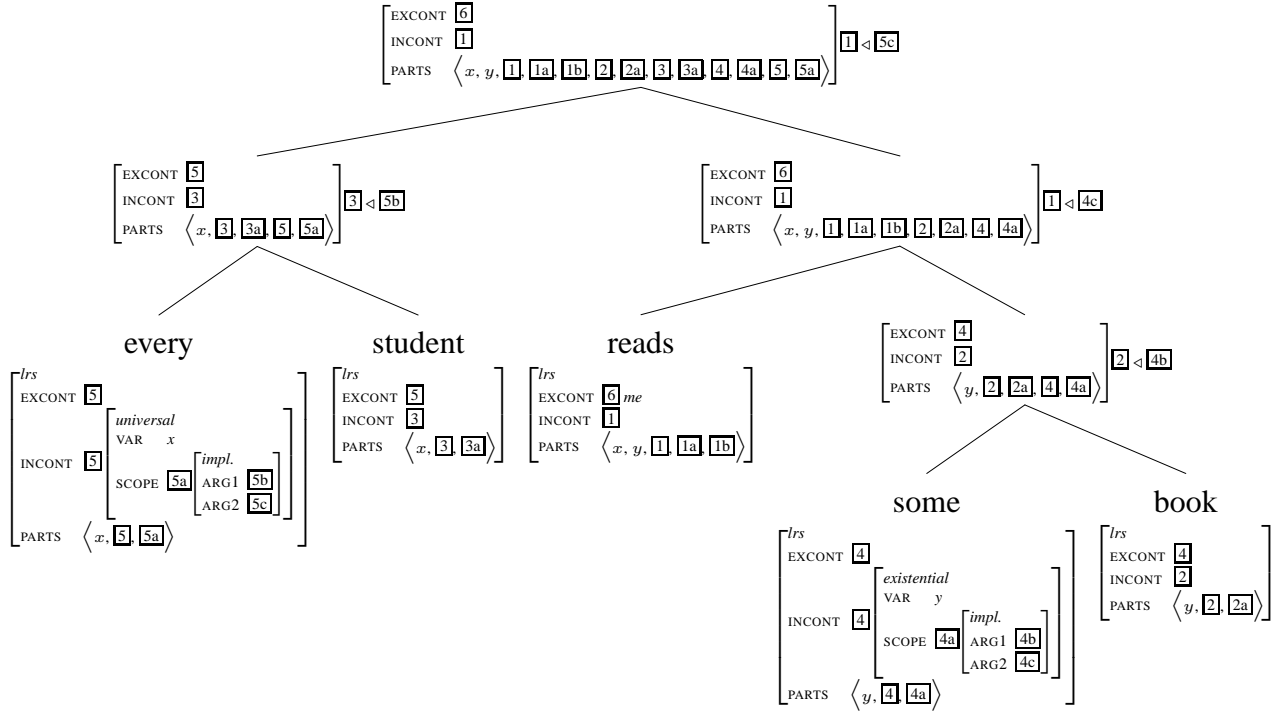


FIG. 1 – Every student reads some book

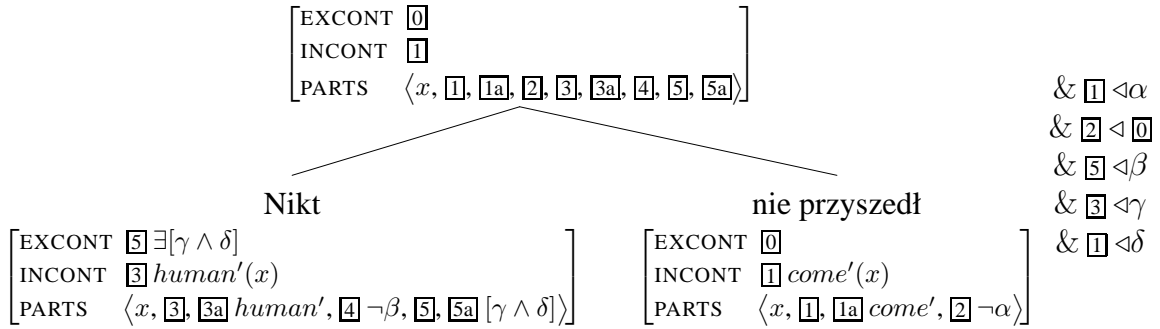


FIG. 2 – Nikt nie przyszedł / Personne n'est venu

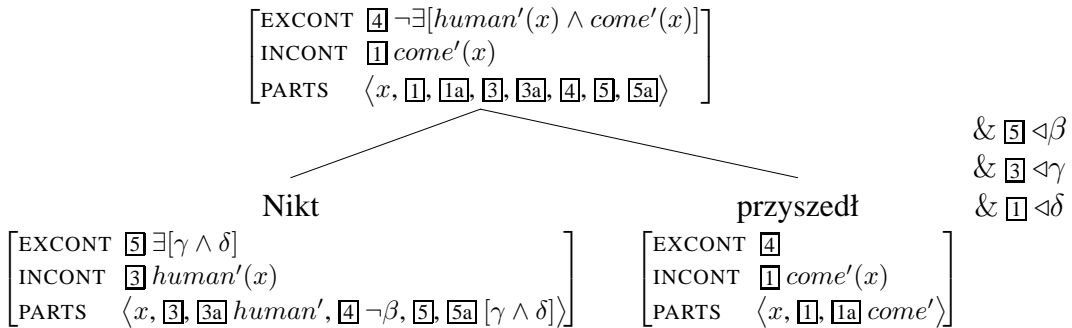


FIG. 3 – Nikt przyszedł / Personne est venu