



**HAL**  
open science

## Quantifier Scoping and Semantic Preferences

Davide Catta, Mehdi Mirzapour

► **To cite this version:**

Davide Catta, Mehdi Mirzapour. Quantifier Scoping and Semantic Preferences. CONLI: Computing Natural Language Inference, Sep 2017, Montpellier, France. lirmm-01916224

**HAL Id: lirmm-01916224**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01916224>**

Submitted on 8 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Quantifier Scoping and Semantic Preferences\*

Davide Catta  
LIRMM, Université de Montpellier  
davide.catta@lirmm.fr

Mehdi Mirzapour  
LIRMM, Université de Montpellier  
mehdi.mirzapour@lirmm.fr

## Abstract

This paper addresses the problem of ranking valid semantic readings of a given multiple-quantifier sentence. The proposed model is based on extending Morrill’s account on the syntactic complexity profiling of the categorial proof-nets. The extension uses some techniques that are inspired by Hilbert’s epsilon operator and the syntactic structure constraints. Our preferential model on different semantic readings can potentially be integrated into natural language inference systems.

## 1 Introduction

The ultimate goal of this research is to rank valid semantic meanings of a given multiple-quantifier sentence that can be gained from both syntactic and semantic representations; and for the time being, without taking into account other possible factors such as common sense and lexical knowledge.<sup>1</sup>

The described problem is not purely a theoretic one; it is raised with the intention of application in the project e-fran AREN (ARgumentation Et Numrique) for the automated analysis of online debates by high-school pupils. The syntactic formalism that is used in the project is categorial grammars which are suitable for deriving meanings as the logical formulas. Grail (Moot (2017, 2012)), an automated theorem prover is used for parsing. It relies on a chart-based system close to categorial natural deduction which can rather smoothly be transformed to proof-nets.<sup>2</sup> The problem of ranking the possible readings can be tackled if we employ the existing techniques that use proof-nets for measuring the complexity of readings (Morrill (2000); Johnson (1998)). We are specifically interested in Morrill’s approach, but as we will see adopting this technique to our problem is not as straightforward as it seems. Our goal in this research is to fix these issues and improve Morrill’s technique for the representing and ranking semantic ambiguities for multiple-quantifier sentences.

Since sentences with quantifier scope ambiguities often have two (or more) non-equivalent semantic readings, capturing these phenomena with some plausible ranking techniques can potentially be useful for natural language inference systems. It is worth mentioning that different readings allow different inference paths and as it is shown in some studies (Durand-Guerrier (2016)) we can trace back the root cause of some *invalid inferences* in the misinterpretation of quantifiers as when occurs in quantifier scoping phase.

---

\*This work is partially supported by the e-fran project AREN (ARgumentation Et Numrique) for the automated analysis of online debates by high-school pupils.

<sup>1</sup>For a historical survey on this topic see chapter “The Natural History of Scope” in Steedman (2012).

<sup>2</sup>Two important features of the AREN project: (i) since, the data are provided in written language by the high-school pupils, we do not need to model intonation and prosody as a cue for disambiguation. (ii) there are, on average, 20 words in each sentence with almost two hundred category types, and the replies of the users happen with some delays. This restriction makes our solution scalable for this specific project that does not demand any big data analysis. Hence, it does not raise the issues of real-time processing and complexity.

## 2 Advantages and Drawbacks of the Morrill’s Approach

A proposal is put forward (Morrill, 2000, p.333) which works very well for a wide variety of performance phenomena such as garden-pathing, the unacceptability of center embedding and preference for lower attachment. It works in Lambek categorial grammar framework (Lambek (1958), (Moot and Retoré, 2012, Chapter 2)). The measurement uses syntactic analysis represented in the proof-net structure and gains the complexity profile by counting the number of unresolved dependencies of each word. This measurement is supposed to represent, in a numerical way, the course of memory load in optimal incremental processing. Formally speaking, we can rewrite Morrill’s definition as follows: let  $s = w_1, \dots, w_n$  be a given sequence of words with  $n$  size, and let  $R_1, \dots, R_m$  be  $m$  number of categorial proof nets that represent all the possible (cut-free) semantic readings of  $s$ . We can define a function  $f_{cm}$  from  $P = \{R_1, \dots, R_m\}$  to  $\mathbb{R}$  that maps  $R_i$  to the inverse of the sum of unresolved valencies before and after each root in  $R_i$ . We can say  $R_i >_m R_j$  (read as  $R_i$  is Morrill-preferred to  $R_j$ ) iff  $f_{cm}(R_i) > f_{cm}(R_j)$ .

There are two general objections against this approach. The first objection comes back to the choice of Lambek categorial grammar which simply can not derive all valid semantic readings. We address this objection in section (3.2) by providing a complexity measuring technique that works only on the valid readings, and not over-generated readings that exists in the storage-based techniques (Cooper (1983), Keller (1988)). The second objection, which is more important than the first one, is that Morrill’s approach simply fails in some examples such as utterances that have sentence-modifier adverbials, nested sentences or direct speech. We can consider the following sentences:

- (1) Everyone repairs something expertly.<sup>3</sup>
- (2) ”Someone loves everyone”, said John.

Figure 1: Categorical Proof-nets and Complexity Profile for Example (1)

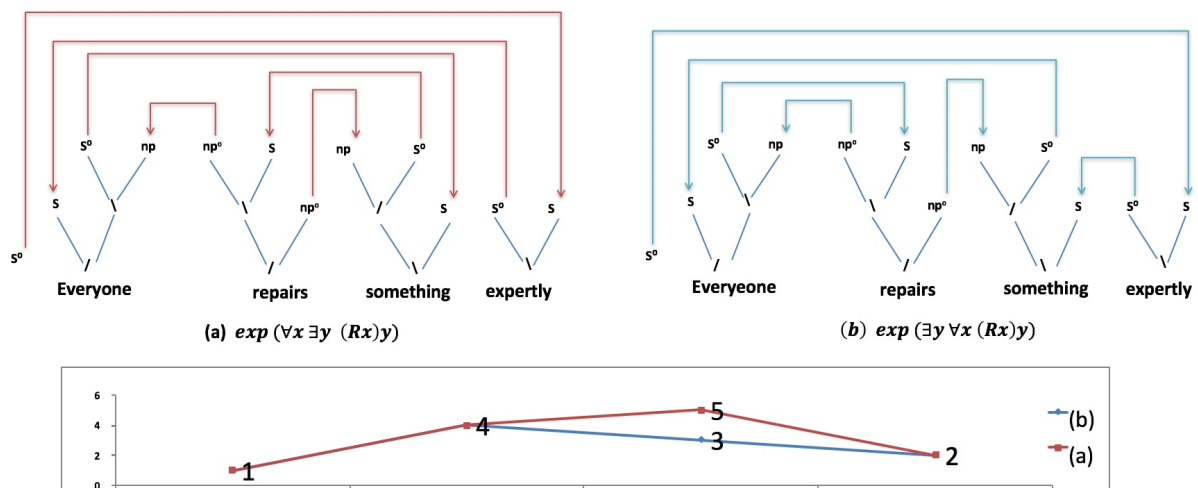


Figure (1) illustrates two corresponding proof-nets and complexity profiles for the example (1)– the reader can verify the example (2) for herself. As it is shown, the Morrill’s criteria for choosing the left-to-right preference fails, since the reading (a) which is supposed to have lower complexity is predicted

<sup>3</sup>Thanks to one of the reviewers that pointed up the distinction between VP-oriented adjunct and Clause-oriented adjunct as it is indicated in (Huddleston et al., 2002, p.575-578). Examples such as *someone loves everyone*, *unfortunately* belongs to clause-oriented adjunct category. In this kind of examples, end positions for adverbs are strongly disfavoured unless there is a prosodic detachment. Our example in the paper is a kind of VP-oriented adjuncts and it prefers end position where prosodic detachment is not normal. Taken this grammatical note into account, we can stress that our example is correct without the need to use a comma before the adverb *expertly*.

wrongly by Morrill’s complexity profiling. This objection which shows that Morrill’s criteria are context sensitive would naturally lead us to demand a new proposal for multiple-quantifier sentences.

### 3 A New Proposal

Although it is true that Morrill’s proposal fails in some cases related to the scoped readings, it is also true that it works rather well for other syntactic phenomena. Considering this fact, our strategy is to keep Morrill’s criteria unchanged for those phenomena that work well and add a new procedure just for evaluating the complexity of quantification. We will see how we can filter out the complexity which can be raised by quantifiers. Formally, given two syntactic structures  $S_1, S_2$  for a given sentence, the complexity can be evaluated in the following way:

$$C(S_1) > C(S_2) \iff (a_1, b_1) > (a_2, b_2)$$

where  $a_1, a_2$  are the results of Morrill’s criteria on  $S_1$  and  $S_2$ , and  $b_1$  and  $b_2$  are the results of our proposal regarding quantifier order measurement (given in section 3.2). Notice that we are using lexicographical order, namely,  $(a_1, b_1) > (a_2, b_2)$  iff  $(a_1 > a_2)$  or  $(a_1 = a_2$  and  $b_1 > b_2)$ . Broadly speaking, we will take the following steps:

1. Using in situ quantifier type assignment for construction the categorial proof-nets in order to filter out the quantifier effects in complexity profiling phase. (Described in sub-section 3.1)
2. Measuring syntactic complexity profiles for the previous step.
3. Measuring quantifier distance for different valid readings. (Described in sub-section 3.2)
4. Introducing the preference relations from the last two steps.

#### 3.1 In situ (=Overbinding) Quantification

Hilbert’s  $\epsilon$ -calculus (Hilbert (1922)) is receiving a renewed interest. In particular, the application of  $\epsilon$ -calculus to linguistics is becoming appreciated<sup>4</sup>. Our strategy for neutralizing quantifier effects in our complexity measurement on syntactic proof-nets can take place by exploiting Hilbert’s  $\epsilon$ -calculus in our semantic recipes.

In Hilbert’s epsilon aside from the usual terms of first order logic, we have the  $\epsilon$  and  $\tau$  terms. In particular if  $A$  is a formula and  $x$  is a variable then  $\epsilon_x A$  ( $\tau_x A$ ) are terms where all the occurrences of  $x$  in  $A$  are bound by  $\epsilon$  ( $\tau$ ). What is interesting about this terms is that they express existential and universal quantification. In particular,  $\epsilon_x A$  is the generic existential element and  $\tau_x A$  is the generic universal element. The proper type for  $\epsilon$  and  $\tau$  is  $(e \rightarrow t) \rightarrow e$ . If we want to translate the sentence *All man are brave*, using the  $\tau$ -binder, we can rewrite it as *brave*( $\tau_x man(x)$ ). As we can see an  $\epsilon$  expressions can take scope over the entire sentence even if its occurrence is nested in the parsed tree.

We can observe that the formula *ate*( $\epsilon \lambda y.(pizza\ y), \tau \lambda x.(child\ x)$ ) is a proper translation for the sentence *Every children ate a pizza* in Hilbert epsilon fashion. This formula does not correspond to any usual logical formula and it is very similar to underspecified representation. This underspecified representation which properly corresponds to syntax tree can help us to filter out the effect of quantifiers for the sentences in the complexity measurement phase of the categorial proof-nets. We do not have to perform the quantifier type raising in the syntactic phase in order to properly represent quantification. So, we can keep the syntactic type of quantifiers as being *np/n*. As discussed, typed Hilbert’s epsilon

<sup>4</sup>For an introductory explanation on Hilbert epsilon see Chatzikyriakidis et al. (2017) and (Retoré, 2014, p.218-221).

(Retoré, 2014, p.218-221) suggests some properties that does not exists in other underspecified representations (Copestake et al. (2005), Cooper (1983) and Steedman (1999)). The first projection, namely  $a$  in our complexity pair representation  $(a, b)$ , can be introduced by the following procedure:

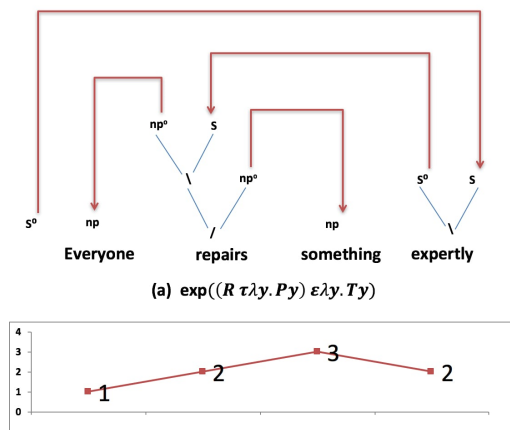
1. Define  $np/n$  syntactic categories to all determiners and quantifiers in our lexicon.
2. Define proper Hilbert epsilon semantical representation for the quantifiers in our lexical recipes.
3. Construct the categorial proof-nets for the sentence with the categorial assignments in our lexicon and plug the lexical recipes to each word.
4. Calculate the complexity measurement profiles of all the valid constructed categorial proof-nets gained in the previous step. This is the first projection of the pairs in our preference semantic model.

### 3.2 Quantifiers Order Measurement

Now, we can introduce the second projection, namely  $b$ , in our complexity pair representation  $(a, b)$ :

1. Given the logical formula which corresponds to the left-to-right reading of the sentence; add an index from 1 to  $n$  to each quantifier from left to right obtaining  $Q_1, Q_2, \dots, Q_{n-1}, Q_n$  call this formula  $\Phi$ .
2. By the procedure introduced in (Hobbs and Shieber, 1987, p.49-53), derive all the valid quantifier (scope) readings of the sentence.
3. Let  $\xi_1, \dots, \xi_m$  be rewritten formulas obtained from the previous step.
4. Calculate for each  $\xi_i$  the penalty of quantifiers re-ordering as  $\frac{1}{\sum_{j=1}^n |j - Pos(Q_j, \xi_i)| + 1}$  in which  $Pos(Q_j, \xi_i)$  is the occurrence position of the quantifier  $Q_j$  in  $\xi_i$  counted from left to right and incremented from number one.
5. The preference measurement on the quantifier ordering of each valid logical readings is gained from applying the previous step to all  $\xi_1, \dots, \xi_m$ .

Figure 2: New Procedure for Example (1)



Let us re-calculate the complexity profile of the example (1) with our new proposal. By the procedure provided in (3.1) we can have the proof-net as it is shown in the figure (2). By assigning the

semantic recipes and having the associated  $\lambda$ -term to the proof-net we will have the unspecified semantic representation that naturally corresponds to the linguistic syntactic structure<sup>5</sup>. As the effect of quantifiers is neutralized we can see that the first projection of our pair is fixed by the number 8. So, we can calculate the second projection namely the score for each reading using the procedure in (3.2), thus we have:

Reading (a) is  $\frac{1}{(|1-1| + |2-2|) + 1} = 1$ , and reading (b) is  $\frac{1}{(|1-2| + |2-1|) + 1} = \frac{1}{3}$ . Thus, we have  $(8, 1) > (8, 0.3)$ , and this shows that reading (a) is preferred to reading (b). All in all, the new proposal truly predicts the quantifier left-to-right scoping while the Morrill’s criteria did not.

## 4 Conclusion and Possible Extensions

We have reported an issue provided by some counter examples on Morrill’s complexity profiling approach. We showed how exploiting Hilbert epsilon can neutralize the quantifier effect and let us introduce a new procedure for measuring the complexity of quantifier scoping. We can extend our on-going project in the following directions: (i) ideally, there should be a dataset annotated for human-preferred readings of various naturally occurring utterances. In the absence of this, we are going to create a test suite illustrating the various phenomena that interact with quantifier scope preferences, and then to evaluate and extend it with the proposed method over that test-suite. This experiment would pave the way to the extension of our model by integrating other aspects such as common-sense knowledge and lexical semantics. (ii) we can exploit Montagovian Generative Lexicon framework (Retoré (2014)) that uses multi-sorted logical formulas. This choice would let us apply semantic distance approaches (Nagao (1994)) by using lexico-semantic networks such as Jeux-De-Mot (Lafourcade (2007); Chatzikiyriakidis et al. (2015)).

## 5 Acknowledgement

We would like to express our gratitude to Christian Retoré and Richard Moot for their great support and constructive discussion during this research. We would also like to thank Jean-Philippe Prost and three anonymous reviewers for their insightful comments.

## References

- Chatzikiyriakidis, S., M. Lafourcade, L. Ramadier, and M. Zarrouk (2015). Type theories and lexical networks: Using serious games as the basis for multi-sorted typed systems. In *ESSLLI: European Summer School in Logic, Language and Information*.
- Chatzikiyriakidis, S., F. Pasquali, and C. Retoré (2017). From logical and linguistic generics to hilberts tau and epsilon quantifiers. *IfCoLog Journal of Logics and their Applications* 4(2), 231–255.
- Cooper, R. (1983). Quantification and syntactic theory, reidel, dordrecht. *CooperQuantification and Syntactic Theory*1983.

---

<sup>5</sup>An anonymous reviewer has pointed out an objection: our proposed epsilon-style representation is not generally valid because of the equivalence  $\tau_x Px \equiv \epsilon x \neg Px$ , and the fact that the epsilon is intuitively interpreted as a choice function. This will affect us when non-P ( $\neg P$ ) is chosen by the choice function. This objection will not cause any problem for our proposal since, firstly, the equivalence is not valid in intuitionistic logic; secondly, even if we use classical epsilon-calculus, it is debatable that  $\epsilon$  - terms can be interpreted as choice functions. In fact, choice functions are only defined on non-empty domains. If we consider a domain of interpretation, where each element is in  $P$ , we can not interpret  $\epsilon$  as a choice function given that there is no non-empty subset of  $\neg P$  where the choice function can be defined. To our knowledge, there is no completely satisfactory model of the epsilon-calculus aside from the indexed version as in Leib (2017).

- Copestake, A., D. Flickinger, C. Pollard, and I. A. Sag (2005). Minimal recursion semantics: An introduction. *Research on Language and Computation* 3(2-3), 281–332.
- Durand-Guerrier, V. (2016). Négation et quantification dans la classe de mathématiques.
- Hilbert, D. (1922). Die logischen grundlagen der mathematik. *Mathematische Annalen* 88(1), 151–165.
- Hobbs, J. R. and S. M. Shieber (1987). An algorithm for generating quantifier scopings. *Computational Linguistics* 13(1-2), 47–63.
- Huddleston, R., G. K. Pullum, et al. (2002). The cambridge grammar of english. *Language*. Cambridge: Cambridge University Press, 1–23.
- Johnson, M. (1998). Proof nets and the complexity of processing center embedded constructions. *Journal of Logic, Language and Information* 7(4), 433–447.
- Keller, W. R. (1988). Nested cooper storage: The proper treatment of quantification in ordinary noun phrases. In *Natural language parsing and linguistic theories*, pp. 432–447. Springer.
- Lafourcade, M. (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *7th International Symposium on Natural Language Processing (SNLP'07)*.
- Lambek, J. (1958). The mathematics of sentence structure. *The American Mathematical Monthly* 65(3), 154–170.
- Leiß, H. (2017). On equality of contexts and completeness of the indexed epsilon-calculus. *IfCoLog Journal of Logics and their Applications* 4(2), 347–366.
- Moot, R. (2012). Wide-coverage semantics for spatio-temporal reasoning. *Traitement Automatique des Langues* 53(2), 115–142.
- Moot, R. (2017). The Grail theorem prover: Type theory for syntax and semantics. In Z. Luo and S. Chatzikyriakidis (Eds.), *Modern Perspectives in Type Theoretical Semantics*. Springer. to appear.
- Moot, R. and C. Retoré (2012). *The logic of categorial grammars: a deductive account of natural language syntax and semantics*, Volume 6850. Springer.
- Morrill, G. (2000). Incremental processing and acceptability. *Computational linguistics* 26(3), 319–338.
- Nagao, K. (1994). A preferential constraint satisfaction technique for natural language analysis. *IEICE TRANSACTIONS on Information and Systems* 77(2), 161–170.
- Retoré, C. (2014). The montagovian generative lexicon lambda tyn: a type theoretical framework for natural language semantics. In *19th international conference on types for proofs and programs (TYPES 2013)*, Volume 26, pp. 202–229.
- Steedman, M. (1999). Alternating quantifier scope in ccg. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 301–308. Association for Computational Linguistics.
- Steedman, M. (2012). *Taking scope: The natural semantics of quantifiers*. MIT Press.