Proceedings of the ESSLLI 2016 Student Session

28th European Summer School in Logic, Language & Information
August 15-26, 2016, Bozen-Bolzano, Italy

Marisa Köllner
Ramon Ziai
(Editors)

Preface

These proceedings contain the papers presented at the Student Session of the 28th European Summer School in Logic, Language and Information (ESSLLI 2016), taking place at the Free University of Bozen-Bolzano from August 15th to 26th, 2016. The Student Session is part of the ESSLLI tradition and has been organized for the twentyfirst time this year. It is an excellent venue for students to present their work on a diverse range of topics at the interface of logic, language and information and to receive valuable feedback from renowned experts in their respective field. The Student Session this year attracted submissions from 18 different countries and has thereby become a great event for students all over Europe and beyond. As in previous years, we have received high-quality submission which made it very hard for us and the reviewers to make out acceptance decisions. We received 43 submissions, 31 of which were submitted for oral presentations, and 12 of which were submitted for poster presentations. At the Student Session, 16 of these submissions were presented as talks and 8 submissions were presented in form of a poster. Due to special request of three authors, their papers were not included in the online proceedings.

We would like to thank each of the co-chairs for all their invaluable help in the reviewing process and organization of the Student Session. Without them, the Student Session would not have been able to take place. Additionally, we would like to thank the area experts for their help in the reviewing process and their support of the co-chairs. We would also like to thank the ESSLLI Organizing Committee, for organizing the entire summer school, and catering to all our needs. Thanks go to the chairs of the previous Student Sessions, in particular to Ronald de Haan, Philip Schulz and Miriam Kaeshammer, for providing us with many of the materials from the previous years and for their advice. As in previous years, Springer-Verlag has generously offered prizes for the Best Paper and Best Poster Award, and for this we are very grateful.
Most importantly, thanks to all those who submitted papers, for they are the ones that make the Student Session an exciting and great event. We encourage them to keep doing such excellent work.

August 2016

Marisa Koellner & Ramon Ziai
Chairs of the ESSLLI 2016 Student Session

Organization

Program Committee

Chairs

Marisa Köllner (*Universität Tübingen*)
Ramon Ziai (*Universität Tübingen*)

Language & Computation co-chairs

Enrico Santus (*Hong Kong Polytechnic University)*
Vered Shwartz (*Bar-Ilan University)*

Logic & Computation co-chairs

Jon-Hael Brenas *(Grenoble Informatics Laboratory)*
Shqiponja Ahmetaj *(Technische Universität Wien)*

Logic & Language co-chairs

Davis Ozlos *(Fribourg University)*
Karoliina Lohiniva *(University of Geneva)*

Area Experts

Language & Computation

Alessandro Lenci *(University of Pisa)*
Ido Dagan *(Bar-Ilan University)*

Logic & Computation

Magdalena Ortiz *(TU Wien)*

Logic & Language

Carla Umbach *(ZAS Berlin/University of Cologne)*
Yoad Winter *(Utrecht University)*

Table of Contents

Language and Computation

Logic & Computation

Logic & Language

Using Named Entities to Discover Heterogeneous Events on Twitter

Amosse Edouard

Université Côte dAzur, Inria, CNRS, I3S, France
amosse.edouard@unice.fr

Abstract. Social media sites such as Twitter[1] and Facebook[2] have emerged as powerful means of communication that allow people to exchange information about their daily activities, latest news or real-world events. Beside social interactions among users, social medias are expected to provide added value services in a variety of domains (e.g sentiment and trend analysis, event detection). Detecting events on social medias poses new challenges due to the sparsity and the informal nature of social media posts. One of the main challenges in detecting events in social media is to differentiate messages concerning events from the others. To face this challenge, we propose to take advantage of the knowledge that can be extracted from the Linked Opened Data (e.g. DBpedia) to enrich the short textual messages with contextual information brought by the presence of named entities. We evaluate our approach on two gold-standard datasets and the preliminary results show that exploiting the ontological categories of the named entities has a positive impact on classification.

Keywords: Event Detection, NLP, Supervised Classification, Named Entities

1 Problem Statement and Motivation

The analysis of social media streams, particularly Twitter, has gained a lot of interest, both within the academic and business communities. The capability to understand and analyse the stream of messages on Twitter is an effective way to monitor what people think [14], which trending topics are emerging [21], and which main events are affecting people's lives. For this reason, several automated ways to track and categorise *events* on Twitter have been proposed in the literature. However, the sheer amount of information contained in tweets makes it a challenge compared to other source of information such as media news. There are two main reasons for this. One is the larger and real time amount of information compared to media news. Second, the characteristics of tweets (e.g. limited to 140 characters long) and specific language. More importantly, contrary to media news, not all information from Twitter is related to events [12].

[1] http://twitter.com/

[2] http://facebook.com/

Recently, detecting events by analyzing tweets has been widely investigated in the field of information retrieval. Although several approaches have focused on the detection and the analysis of large-scale events from Twitter, most of them focus on detecting specific types of events [2, 10, 22]. Such approaches are highly dependant of the target events and mainly rely on specific keywords to filter event-related tweets. Their strong connection to the target event makes them unsuitable for detecting events at Twitter's scale.

In the context of events in social medias, Dou et al. [9] define and event as "*An occurrence causing change in the volume of text data that discusses the associated topic at a specific time. This occurrence is characterized by topic and time, and often associated with entities such as people and location*". This definition, adopted in our work, highlights a strong connection between events in the context of social media and the named entities (NEs) involved in such events (i.e. events' participants, typically persons, organisations and locations). Moreover, NEs can be linked to external knowledge bases (KBs) containing additional semantic information, such as DBpedia[3].

The goal of this paper is to propose an approach to detect and classify real-world events on Twitter by relying on the knowledge that can be extracted from KBs, to enrich tweets with contextual information brought by the presence of the NEs involved in the events.

2 Related Work

Existing works on event detection can be classified into two main categories: *i) Close-domain* : interested in detecting known event type (e.g. earthquakes, flu pandemic or incidents); *ii) Open-Domain*: interested in detecting unknown event. Although our goal is to identify heterogeneous events (i.e. event of unknown types), approaches that target a particular event type are relevant to highlight the challenges in event detection on tweets. In the reminder of this section, we review existing works in both categories.

Close-Domain or Known Event Type. Works of this group mainly focus on monitoring tweets for detecting known events such as earthquakes [22], incidents [3, 23] or social activities [11]. Usually, a set of keywords related to the target event is used to extract relevant tweets from the Twitter stream. Sakaki et al. [22] propose an approach for detecting earthquakes by monitoring Twitter messages. The keywords "earthquake" and "shake" are used to retrieve relevant tweets. To eliminate off-context tweets (e.g. tweets containing "shake hand"), SVM is used as a binary classifier using features related to earthquakes. Attardi et al. [7] use discriminative word embeddings as continuous features for training an SVM classifier with the aim of separating tweets related to natural disasters from the others using words related or indicative of disasters as lexicon.

Twitter is also used to extract additional information about existing events. For example, Ritterman et al. [21] predict swine flu pandemic in 2009. Achrekar

[3] `http://wiki.dbpedia.org/`

et al. [4] use tweets related to the flu as early indicators of influenza-like illness. Abel et al. [2, 3] use semantic linking to filter relevant information from tweets about reported incidents in an emergency broadcasting service. Keywords related to the reported incidents are used to extract relevant tweets. Semantic linking is used to find particular information pieces in the relevant tweets.

To identify tweets related to events, also the event type is used to create event patterns for detecting fine-grained topics [24, 26] or to define labels for training Machine Learning classifiers [5, 22]. Due to their strong connection to the target event, such approaches cannot be applied to event detection at Twitter's scale unless one knows the keywords corresponding to each event in advance (not straightforward).

Open-Domain or Unknown Event Type. Two approaches are mainly used to detect open-domain events on Twitter : Document-Pivot or Feature-Pivot techniques [6]. In the former, documents are clustered on the basis of their textual similarity, while the latter monitors bursty terms (i.e. terms that are observed at an unusual rate) in a collection of documents, where a bursty term is considered as indicator of an event. Petrovic et al. [18] address the First Story Detection task by analysing solely the contents of tweets. Their approach is based on local sensitive hashing, a randomized technique that reduces the time needed to find a nearest neighbor in a vector space. Each new tweet is assigned to the thread that contains the most similar tweets, where similarity is based on cosine similarity. The growth rate of thread is used to eliminate non-event related threads, such that threads that grow fastest are considered as event-related.

Ritter et al. [20] model events on Twitter as a 4-tuple representation including NEs, temporal expressions, event phrases and event type. NEs and temporal expressions are extracted using Twitter specific tools [19] while event phrases are extracted using a supervised method. The system recognizes event triggers as a sequence labeling task using Conditional Random Field; then an unsupervised approach is used to classify the events into topics. In addition, the authors consider the association strength between NEs and temporal expressions to decide whether or not a tweet is related to an event. This assumption restricts the approach to tweets that explicitly contain temporal expressions and NEs.

Most of the existing works on open-domain event detection on Twitter rely on the speed according to which the clusters are growing: clusters that grow faster are considered as event-related [20, 27]. Although this assumption helps in discovering large-scale events [17], it is less suitable for events with a small audience on Twitter. In our work, instead of creating event clusters on the whole Twitter stream, we propose to separate event and non-event tweets in a separated task. Our work is partially inspired by [23] to generalize an event classifier model by replacing the NEs with their semantic categories in ontologies. Finally, we create event clusters using only tweets that are related to events.

3 Research Questions and Working Hypotheses

Our main research questions are: *how to detect open-domain events by monitoring Twitter messages? What is the best approach to build an event detection model that can hold good performance as time passes?* More precisely, we will investigate the following subquestions:

RQ1: Is it possible to use supervised classification to separate event-related from not event-related tweets? What is the impact of in-domain data on classification?
RQ2: How information contained on the Linked Open Data (as knowledge about NEs) can contribute to this task in order to mitigate the effects of overfitting on in-domain data?
RQ3: How can we cluster/categorize events into finer-grained topics?

To address the above questions, we make the following working hypotheses:

H1: Separating event-related from non-event related tweets can contribute in reducing the computational time of event detection algorithms.
H2: The presence of NEs in the content of a post is a good indicator that it is related to an event.
H3: Replacing NEs in tweets by their corresponding category in an ontology can reduce the negative effect of overfitting on a classifier.

4 Proposed Approach

In the following, we describe the approach we propose for detecting heterogeneous events on Twitter (i.e. unknown event types), consisting in three steps:

1. We separate event-related tweets from the rest of the micro-posts by combining techniques from Machine Learning, NLP and LOD.
2. We classify the tweets that are related to events into coarse-level categories as described in the TDT manual [1] including: Science, Armed Conflicts, Politics, Economy, Culture, Sports, Accidents and Miscellaneous.
3. We propose to cluster the tweets in each category into finer-grained topics by grouping similar tweets using a feature-pivot technique.

In the following, we provide a steps-by-step description of the approach.

4.1 Identifying Event-Related Tweets

In previous works [1, 9, 11, 20, 23], events are typically defined according to time, space and agents involved such as locations, persons or organisations, denoted as named entities. For the first and second step, we propose to build a supervised model based on semantic abstraction on the NEs. Our semantic abstraction consists in replacing the NEs cited in tweets by their ontological categories (e.g.

their type in DBpedia) and use the modified content to extract features for training a supervised model. More specifically: we first link the NE mentions in tweets to resources in a KBs (i.e. DBpedia); second, we replace the NEs by their category in the ontology, third, we create a feature vector with the modified content and finally, we use the feature vector to train a supervised model.

Named Entity Recognition, Linking and Replacement We use NERD-ML [25] to perform Named Entity Recognition (NER) and linking, given that [8] demonstrates that NERD-ML performs better on NER on Twitter data than other Twitter-specific NLP tools such as Tweet NLP [19]. SPARQL queries are used to retrieve the categories of the NE in the KB, and we sort the output of the query according to the hierarchy of the ontology. We experiment two NE replacement techniques, namely generic and specific replacement. In the former, NEs are replaced by their most generic category;[4] in the latter, we replace the NEs by their most specific category.[5] Two example outputs of the entity replacement module are reported in Table 1. The rationale behind the replacement of entity mentions with their type is to generalise over single mentions, thus avoiding overfitting in supervised settings.

Original Tweets	**Generic Categories**	**Specific Categories**
Cambodia's ex-King Norodom Sihanouk dead at 89 http://q.gs/2IvJk #FollowBack	[Place] ex-king [Person] die at [number]	[Country] ex-king [Royalty] die at [number]
Amy Winehouse, 27, dies at her London flat http://bit.ly/nD9dy2 #amyWinehouse	[Person], [number], die at her [Place] flat [Person]	[Person], [number], die at her [Settlement] flat [Person]

Table 1. Entity replacement strategies using categories in the DBpedia ontology.

Classification Approach To separate event tweets from non events tweets and to associate a coarse-level category to event-related tweets, we use a supervised method. We consider two strategies: (1) A pipeline: A binary classifier that classifies the tweets into events and non events provides the input to a second model to associate an event category to the tweets. (2) A single multi-class classification: A model trained on 9 classes, including the 8 event categories plus a non event-related class. We plan to experiment both approaches.

4.2 Extracting Fine-grained Event Topics

The third step of our approach is topic detection. Giving a set of tweets label as related to events in the previous tasks, the goal is to detect fine-grained event

[4] i.e. The last category in the hierarchy (excluding the Thing class).

[5] i.e. The first category in the hierarchy.

topic (e.g. The death of Amy Winehouse). Given that the possible event topics in Twitter are unknown in advance [20], we propose to build an unsupervised model exploiting the event categories output by the supervised model.

We create topic clusters in each event category by grouping similar tweets. Due to the sparsity of tweets, we propose to use feature-pivot techniques instead a document-pivot techniques [6]. Instead of considering each word in a tweet as a bursty candidate, we reduce the feature space by considering relationship between NEs and event phrases (i.e. action verbs) as in [28]. Furthermore, using the relationship between NEs and event phrases is useful to separate events sharing a common type into specific event topics (e.g. an accident that occurs in different places or a celebrity involved in different events). Also, we use Wordnet [16] to extract the synsets for the event phrases in order to group similar clusters.

Finally, following the state of the art approaches, emergent event topics are identified by monitoring the growth of each cluster. Since this task is connected to the classification task, all the clusters are related to events; thus, we use the growth rate of the clusters to sort the events according to their popularity in Twitter such that clusters that grow fastest are the most popular.

5 Evaluation Plan

Since we are interested in detecting events on Twitter, we evaluate our approach on two gold-standard datasets of tweets. In the reminder of this section, we present the characteristics of each dataset as well as the evaluation strategies planned for each research question.

5.1 Datasets

For our experiments, we choose two corpora of tweets collected over two distinct periods and cover different and specific events. Both datasets are manually annotated and tweets related to events are annotated either with a coarse-level event category (e.g. Culture) and fine-grained event type (e.g. The death of Amy Winehouse).

The Events 2012 Corpus [15] A total of 120 million tweets were collected from October to November 2012 using the Twitter streaming API,[6] of which 159,952 tweets were labeled as event-related. This corpus contains 506 event types gathered from the Wikipedia Current Event Portal. Amazon Mechanical Turk was used to annotate each tweet with one of such events. Besides, each event was also associated with an event category following the TDT annotation manual [1]. Events covered by this dataset include for example the US presidential debate between Barack Obama and Mitt Romney, the US presidential election results or the Chemistry Nobel prize. After removing duplicated tweets and those that are no-longer available, we are left with ∼92 million tweets from which 42,334 tweets related to events.

[6] `https://dev.twitter.com/streaming/overview`

First Story Detection Corpus (FSD) [18] A corpus of 50 million tweets, collected from July 2011 until September 2011. Two experts annotated the tweets related to events with one out of 27 event types extracted from the Wikipedia Current Event Portal; agreements between the annotators using Cohen's kappa was 0.65. In total, 3,035 tweets were labeled as related to events and annotated with a corresponding event topic (e.g. 'death of Amy Winehouse', 'earthquake in Virginia' or 'plane crash of the Russian hockey team'). After removing tweets that are no longer available, we are left with ~31 million tweets from which 2,250 are related to events.

Contrary to the Event 2012 corpus, the events in the FSD corpus are not associated with event categories. Therefore, in order to merge the two corpora in a single dataset for our experiments, we extended the FSD corpus by labelling each event topic with one of the event categories of the Event 2012 corpus [15]. The task was manually performed by three annotators: the labels were first assigned independently, and then adjudicated by majority vote in case of disagreements.[7] Agreement between the three annotators, measured using Krippendorffs alpha coefficient, was (alpha = 0.758). Table 2 shows the number of tweets in each corpus divided into categories.

Event Category	**Event 2012**	**FSD**
Arts	2589	710
Attacks	7079	56
Politics	16383	58
Sports	8812	0
Economy	2881	342
Science	1537	296
Accidents	2479	778
Miscellaneous	574	10
Total	42334	2250

Table 2. Tweets in each event category

5.2 Evaluation Strategies for the First and Second Research Questions

To demonstrate the importance of NLP in detecting open-domain events on Twitter, we compare our NLP-based approach against a baseline which does not make use of NLP nor LOD. On the other hand, our evaluation aims to prove that it is feasible to use supervised machine learning to separate event related tweets and non event related tweets.

Since both corpora contain much more non-event related than event related tweets, resulting in a very skewed class distribution, we reduced the number of

[7] The Web interface used for annotation is available at `http://www.i3s.unice.fr/~edouard/events/agreements.html`

negative instances by randomly selecting a sample of non event-related tweets. The final amount of tweets in the two datasets is reported in Table 3.

We consider two evaluation settings namely, Setting 1 and Setting 2. In Setting 1, we evaluate the model using 10-fold cross validation only with the Event 2012 corpus. In Setting 2, we use the Event 2012 as training set and the FSD corpus as test set. In our work, we focus on Setting 2 (i.e. train the model on a corpus and test it on the other); however, Setting 1 is useful to understand the effect of only in-domain data on the output of supervised method.

	Event-related	**Non event-related**	**Total**
Event 2012	42,334	48,239	90,573
FSD	2,250	3,040	5,290

Table 3. Total number of tweets per dataset

5.3 Evaluation Strategies for the Third Research Question

To evaluate our approach for detecting event topic, we will use the event topics and categories from the datasets described in Section 5.1 as ground truth. Our evaluation strategy is two fold: (1) We evaluate the ability of our approach to determine the correct event topics and (2) we compare the summary of the topics with the summary of each events in the datasets.

6 Experimental setup

We have carried out some experiments to evaluate our approach for separating tweets related to events from the rest of tweets. We compare the obtained results against a simple baseline which does not make use of NLP nor LOD.

Before training the classifiers, we further clean up the datasets to remove Twitter-specific features such as URLs, user mentions, emoticons and duplicate tweets. We also perform standard pre-processing such as stop words removal and stemming. We employ character sequence n-gram features [13] and model the feature vector using bag-of-words weighted by TF-IDF. We consider the NE replacement strategies described in Section 4. Tweets related to events are considered as positive instances and tweets not related to events as negative instances (See Table 3). Weka[8] was used to train both Naive Bayes and SVM classifiers. Tables 4 and 5 report on the results obtained on Settings 1 and 2.

6.1 Preliminary Results

Table 4 reports on the results obtained on Setting 1 (i.e training and test on the Event 2012 corpus). The baseline outperforms our method in precision and recall

[8] http://www.cs.waikato.ac.nz/ml/weka/

for both NB and SVM classifiers. These results are similar to those obtained by [11], that highlights that training and testing on the same datasets bring to higher performances than training and testing on different datasets, due to overfitting (given the similarity between training and testing instances).

We also conduct a set of experiments on Setting 2 (i.e we train on the Event 2012 dataset and test on the FSD dataset) and reports the results in Table 5. With this configuration, the best performing method is obtained when the NEs are replaced by their most generic category in the DBpedia ontology outperforming the baseline. As expected, both methods obtain lower performance with respect to Setting 1. Nevertheless, while our method yields a drop of 0.028in f-measure, the baseline yields a drop of 0.167. These preliminary results show that using the ontological categories of the NEs, we mitigate the impact of overfitting on our supervised model.

	Naive Bayes			SVM		
Approach	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**
dbp:generic.	0.897	0.896	0.896	0.906	0.903	0.904
dbp:specific.	0.871	0.869	0.870	0.887	0.884	0.885
Baseline	0.919	0.918	0.918	**0.951**	**0.949**	**0.950**

Table 4. Evaluation on setting 1: A NB and an SVM classifier are trained and tested with tweets from the Event 2012 dataset using 10-fold cross-validation.

	Naive Bayes			SVM		
Approach	**Prec.**	**Rec.**	**F1**	**Prec.**	**Rec.**	**F1**
dbp:specific.	0.811	0.809	0.810	**0.879**	**0.875**	**0.876**
dbp:generic.	0.810	0.809	0.809	0.873	0.873	0.873
Baseline	0.783	0.784	0.783	0.789	0.739	0.763

Table 5. Evaluation on setting 2: A NB and an SVM classifier are trained with tweets from the Event 2012 dataset and tested with tweets from the FSD dataset.

7 Discussion and Future Work

This paper presents an approach to address the problem of detecting open-domain events on Twitter, together with the obtained preliminary results. The underlying idea relies on the relation between events and NEs involved in such events, but also on the use of both NLP and LOD to build a supervised method to detect tweets related to events, and to classify them into event categories. We are currently investigating the use of the output of the supervised model as input for a clustering algorithm to detect finer-grained event topics within each category.

We found that the replacement of the NEs in tweets by their associated concepts in the DBpedia ontology has proved to be efficient in reducing the negative effect of overfitting. Our preliminary results show that the proposed approach holds higher precision and recall compared to a baseline when the training and test sets are different. For future works we plan to improve the approach by considering NEs categories from other ontologies such as Yago; we also plan to experiment other classifiers such as Neural Network.

Acknowledgments. The author would like to thank Dr. Elena Cabrio, Dr. Sara Tonelli and Prof. Nhan Le-Than for supervising the development of the research plan, their writing tips and the review of the drafts of this paper. I am also grateful to the reviewers for their valuable and constructive feedback.

References

1. TDT 2014: Annotation Manual. `https://catalog.ldc.upenn.edu/docs/LDC2006T19/TDT2004V1.2.pdf` (2014), [Online; accessed 03-March-2016]
2. Abel, F., Hauff, C., Houben, G.J., Stronkman, R., Tao, K.: Semantics+ filtering+ search= twitcident. exploring information in social web streams. In: Proceedings of the 23rd ACM conference on Hypertext and social media. pp. 285–294. ACM (2012)
3. Abel, F., Hauff, C., Houben, G.J., Stronkman, R., Tao, K.: Twitcident: fighting fire with information from social web streams. In: Proceedings of the 21st international conference companion on World Wide Web. pp. 305–308. ACM (2012)
4. Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.H., Liu, B.: Predicting flu trends using twitter data. In: Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on. pp. 702–707. IEEE (2011)
5. Anantharam, P., Barnaghi, P., Thirunarayan, K., Sheth, A.: Extracting city traffic events from social streams. ACM Transactions on Intelligent Systems and Technology 9(4) (2014)
6. Atefeh, F., Khreich, W.: A survey of techniques for event detection in twitter. Computational Intelligence 31(1), 132–164 (2015)
7. Attardi, G., Gorrieri, L., Miaschi, A., Petrolito, R.: Deep learning for social sensing from tweets. In: Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015. p. 20. Accademia University Press (2015)
8. Derczynski, L., Maynard, D., Rizzo, G., van Erp, M., Gorrell, G., Troncy, R., Petrak, J., Bontcheva, K.: Analysis of named entity recognition and linking for tweets. Information Processing & Management 51(2), 32–49 (2015)
9. Dou, W., Wang, K., Ribarsky, W., Zhou, M.: Event detection in social media data. In: IEEE VisWeek Workshop on Interactive Visual Text Analytics-Task Driven Analytics of Social Media Content. pp. 971–980 (2012)
10. Earle, P.S., Bowden, D.C., Guy, M.: Twitter earthquake detection: earthquake monitoring in a social world. Annals of Geophysics 54(6) (2012)
11. Ilina, E., Hauff, C., Celik, I., Abel, F., Houben, G.J.: Social event detection on twitter. In: Web Engineering, pp. 169–176. Springer (2012)
12. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. pp. 56–65. ACM (2007)

13. Kanaris, I., Kanaris, K., Houvardas, I., Stamatatos, E.: Words versus character n-grams for anti-spam filtering. International Journal on Artificial Intelligence Tools 16(06), 1047–1067 (2007)
14. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. Business horizons 53(1), 59–68 (2010)
15. McMinn, A.J., Moshfeghi, Y., Jose, J.M.: Building a large-scale corpus for evaluating event detection on twitter. In: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. pp. 409–418. ACM (2013)
16. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM 38(11), 39–41 (1995)
17. Osborne, M., Petrovic, S., McCreadie, R., Macdonald, C., Ounis, I.: Bieber no more: First story detection using twitter and wikipedia. In: SIGIR 2012 Workshop on Time-aware Information Access (2012)
18. Petrović, S., Osborne, M., Lavrenko, V.: Streaming first story detection with application to twitter. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp. 181–189. Association for Computational Linguistics (2010)
19. Ritter, A., Clark, S., Etzioni, O., et al.: Named entity recognition in tweets: an experimental study. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 1524–1534. Association for Computational Linguistics (2011)
20. Ritter, A., Etzioni, O., Clark, S., et al.: Open domain event extraction from twitter. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1104–1112. ACM (2012)
21. Ritterman, J., Osborne, M., Klein, E.: Using prediction markets and twitter to predict a swine flu pandemic. In: 1st international workshop on mining social media. vol. 9, pp. 9–17. ac. uk/miles/papers/swine09. pdf (accessed 26 August 2015) (2009)
22. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. pp. 851–860. ACM (2010)
23. Schulz, A., Janssen, F.: What is good for one city may not be good for another one: Evaluating generalization for tweet classification based on semantic abstraction. In: Proceedings of the Fifth International Conference on Semantics for Smarter Cities-Volume 1280. pp. 53–67. CEUR-WS. org (2014)
24. Tanev, H., Piskorski, J., Atkinson, M.: Real-time news event extraction for global crisis monitoring. In: Natural Language and Information Systems, pp. 207–218. Springer (2008)
25. Van Erp, M., Rizzo, G., Troncy, R.: Learning with the web: Spotting named entities on the intersection of nerd and machine learning. In: # MSM. pp. 27–30. Citeseer (2013)
26. Wang, X., Gerber, M.S., Brown, D.E.: Automatic crime prediction using events extracted from twitter posts. In: Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 231–238. Springer (2012)
27. Xie, W., Zhu, F., Jiang, J., Lim, E.P., Wang, K.: Topicsketch: Real-time bursty topic detection from twitter. In: Data Mining (ICDM), 2013 IEEE 13th International Conference on. pp. 837–846. IEEE (2013)
28. Xu, W., Grishman, R., Meyers, A., Ritter, A.: A preliminary study of tweet summarization using information extraction. NAACL 2013 p. 20 (2013)

Modelling Subordinate Conjunctions in STAG: A Discourse Perspective

Timothée Bernard

Université Paris Diderot - ALPAGE
timothee.bernard@inria.fr

Abstract Among the discourse connectives – lexical items conveying discourse relations – are the subordinate conjunctions (SubConjs), like *because*, *even if* or *although.* SubConjs have generally been considered a homogeneous category, however previous work has shown they can be divided into two classes according to their syntactic and semantic properties. Similarly, attitude verbs and reporting verbs (AVs) have two different uses in discourse: *evidential* and *intentional.* Drawing from these observations, we propose a STAG model of SubConjs and AVs taking into account both their syntactic and discursive properties.

Keywords: discourse, STAG, subordinate conjunctions, syntax, semantics

1 Introduction

At the discourse level, sentences and propositions are related by *discourse relations* (DRs). DRs can either be *implicit*, i.e semantically inferred, or *explicit*, i.e. lexically signalled. The most common markers of explicit DRs are *discourse connectives*, a group mainly composed of conjunctions, prepositions and adverbs. For instance, (1a) involves an implicit *Consequence* relation and (1b) a *Concession* one explicitly signalled by the *but* connective. Following the conventions of the Penn Discourse TreeBank (PDTB, [13]), we refer to the two arguments of DRs as Arg$_1$ and Arg$_2$ and use italics and bold face respectively to indicate the spans of text for each argument (when such spans of text appears) while the connective lexicalising the relation, if any, is underlined.

(1) a. *Fred was sick.* **He stayed at home**.
    b. *Fred was sick.* But **he came to work**.

We are working with Synchronous Tree Adjoining Grammar (STAG, [15]), a formalism providing a way to describe both syntax and semantics simultaneously, making explicit how they relate with each other. To our knowledge, not much attention has been paid to modelling subordinate conjunctions (SubConjs) from a discourse point of view in STAG. D-STAG [3] analyses discourse with STAG structures, and thus models discourse connectives, but does not take into account the specificities of SubConjs that are discussed here. This is however a necessary step toward both operational discourse parsers and convincing discourse generation systems, and also the purpose of this mainly theoretical work.

The paper is organised as follows. Section 2 presents relevant work highlighting the aspects of SubConjs that we aim to model. In Section 3, we use linguistic tests to determine more precisely the interactions of SubConjs with diverse scope operators. This leads to our STAG proposition presented in Section 4. Section 5 concludes the paper.

2 Relevant Work on Subordinate Conjunctions

2.1 Non-Alignment of Syntactic and Discourse Arguments

It has been shown by a number of works (see [5] for English and [4] for French) that the propositional content of a (syntactic) argument of a discourse connective is not always a (semantic/discourse) argument of the DR lexicalised by the connective. Such mismatches often arise with *attitude verbs* (*to think*, *to know*, etc.) and *reporting verbs* (*to say*, *to deny*, etc.), both grouped here under the label 'AV'. When an AV together with the clause it introduces is an argument of a discourse connective, the AV may (2a) or may not (2b) be included in the discourse argument of the corresponding DR.[1] Following [1], we say the AV is *intentional* in the first case and *evidential* in the second.

(2) a. *Fred went to Peru* although **Sabine thinks he never left Europe**.
b. *Fred went to Peru* although Sabine thinks **he did not go to Lima**.

It is interesting to note that contrarily to *although*, not all discourse connectives can be found with such non-alignments of the syntactic and discourse arguments. It is the case, for instance, of *because*, as illustrated in (3). [9], using the DR hierarchy of the PDTB, observes that a connective lexicalising a COMPARISON or an EXPANSION relation can often be found with a mismatch, whereas it seems impossible for a connective lexicalising a TEMPORAL or a CONTINGENCY relation.

(3) a. *Fred could not come* because **he was not in town**.
b. #Fred could not come because Sabine thinks he was not in town.

2.2 Two Types of Adverbial Clauses

A distinction between two types of adverbial clauses is made by [8]. The first type is the *central adverbial clause* (CAC), which adds an information (time, place, etc.) about the eventuality described in the matrix clause as in (4). The second type is the *peripheral adverbial clause* (PAC), whose function is to structure the discourse (expressing a concession, providing background information, etc.) as in (5).

[1] Why the AV is included or not in Arg_2 is discussed in [5] and [4]. One element is that the AV can be felicitously removed from (2b) while it cannot from (2a). Similarly, an attributing phrase such as *according to Sabine* can be substituted for the AV (with no change in meaning) only in (2b) and not in (2a).

(4) a. *Fred went to Brazil* while **he was a student**.
b. If **it is sunny**, *I'll go outside.*

(5) a. *Fred has been to Brazil* whereas **Sabine has never left Europe**.
b. If **it is sunny**, *why aren't you playing outside?*

Several phenomena are studied in [8] – coordination, ellipsis, ambiguity, and others related to scope, prosody, typography, etc. They all tend to show a greater integration of CACs into their matrix clause than PACs. We will expand on these observations concerning scope phenomena in the next section. It suffices for now to point out that negation and interrogation may scope over CACs but not over PACs. It should also be noted that a CAC cannot contain an epistemic modal if it is *speaker-oriented* (as in (6a) but not in (6c) where *may* is mainly 'John-oriented'), while a PAC can (see 6b). Expressed with the terms of [9]: the syntactic and discourse arguments of a conjunction must be aligned in the case of a CAC, while there can be a mismatch with PACs.

(6) a. #Mary accepted the invitation without hesitation after John may have accepted it. (from [8])
b. *The ferry will be fairly cheap*, while/whereas **the plane** may/will probably **be too expensive**. (from [8])
c. *John is worried* because **he may be ill**.

3 Projection Tests Applied to Subordinate Conjunctions

In order to model SubConjs, we need to understand how they semantically relate to the other components of the sentence. We therefore study them in the context of the five following patterns[2], related to the scope of diverse operators with respect to discourse connectives and their arguments:

Negation: It is not the case that A.
Conditional: If A, B.
Epistemic: It is possible that A.
Interrogation: A?
AV: Sabine thinks that A.

In these patterns, we replace A with 'A_1 CONJ A_2', where CONJ is a SubConj lexicalising a DR R, and try to figure out if Arg_1, Arg_2 and $R(\text{Arg}_1, \text{Arg}_2)$ are logically implied by the resulting sentence. Note that we do not constrain the syntactic structure of the sentence; we do not know *a priori* whether A is made up of a unique constituent or of multiple constituents diversely attached to the rest of the sentence. In this paper, we illustrate the results for *because* and *although* – introducing a CAC and a PAC respectively – although the examples can be extended to many other SubConjs.

[2] These patterns are commonly used to test projection properties [2].

Because (Explanation): Sentences in (7) are the result of applying the negation and the interrogation patterns to an instance of *because* lexicalising *Explanation*.

(7) a. It is not the case that *Fred was absent* because **he was sick**.
b. Was Fred absent because he was sick?

Interpret (7a) in the context of Fred's workplace. A local interpretation of the negation in the matrix clause (scoping only over *Fred was absent*) would be logically incoherent (in the sense that while it is semantically well-formed, it seems impossible or at least very hard to find a situation in which it would be true), so it must have a global interpretation. This is compatible with [8], as *because* specifies some aspect of the event in the matrix clause and thus introduces a CAC. But what do possible continuations tell us about the semantics of (7a)?[3] All sentences in (8) are possible and describe different situations:[4]

- with (8a), neither Arg_1 (*Fred is absent*), nor Arg_2 (*Fred is sick*), nor R (*Explanation*) are true;
- with (8b), only Arg_1 is true;
- with (8c), only Arg_2 is true;
- with (8d), both Arg_1 and Arg_2 are true, but not R.

(8) a. He was there and in perfect shape.
b. He was fine but he missed his train.
c. He still came, even if indisposed.
d. He was indeed ill, but he would have come anyway hadn't he had to take delivery of some important package.

The negation has therefore a global scope over $\text{Arg}_1 \wedge \text{Arg}_2 \wedge R(\text{Arg}_1, \text{Arg}_2)$ and none of these elements are semantically implied by the use of the *because.* We can notice that all these sentences are also acceptable answers for question (7b) when preceded by the negative *no.* So interrogation has the same properties as negation in terms of these conclusions.

In a similar way, the conditional pattern generates (9), which can be coherently followed by any sentence in (10). They illustrate the same four previous configurations: neither Arg_1 nor Arg_2 nor R are true with (10a), only Arg_1 is true with (10b), only Arg_2 with (10c), Arg_1 and Arg_2 true but not R with (10d).

(9) If *Fred got offended* because **Sabine teased him**, then it would mean that he is secretly in love with her.

[3] We accept any kind of continuation, including dialogue, as long as no correction DR nor 'Hey, wait a minute'-style device [14] is involved.

[4] It could be argued that (8a) and (8c) are no acceptable continuations of (7a), implicitly saying that the use of *because* presupposes the truth of Arg_1. However, this interpretation does not seem to be shared by all English speakers, and as it may involve a specific treatment of presupposition, we have preferred to leave it for future research.

(10) a. However, Fred is in very good mood and I know Sabine, she never teases anyone.
b. However, I don’t think she was teasing him.
c. However, he didn’t seemed annoyed at all.
d. However, I don’t think that this is actually the reason.

The same conclusions can be drawn from all other patterns as well: the corresponding operators can have global scope over the whole ‘A$_1$ because A$_2$’ span.

Although (Concession): There has been a lot of discussion since Frege [6] about the semantics of *although.* According to [12], the *Concession* relation is not *at-issue* (which roughly means that while you can *express* it, you cannot *talk about* it; in particular it cannot be easily negated). It is also interesting to remark that this *Concession* is *speaker-oriented*: *although* cannot be used without the speaker committing herself to the relation, even if the connective is under the scope of AVs, which are *presupposition plugs.* Therefore, *although* is often cited as a *conventional implicature* trigger since [7].

First, let us notice that the negation pattern cannot be directly used with a coherent ‘A$_1$ although A$_2$’.[5] To produce a satisfactory utterance such as (11c), it is necessary for the negation to be included in Arg$_1$. The impossibility of sentences such as (11b) shows that with *although* – contrary to what we have just seen with *because* –, a negation in the matrix clause always has a local scope. This is consistent with the analysis in [8], as *although* is not used to precise an event but to give some context for its interpretation and thus introduces a PAC.

(11) a. *Fred ate meat the other day* although **he is a vegetarian**.
b. #It is not the case that Fred ate meat the other day although he is a vegetarian.
c. *It is not the case that Fred refused to eat meat the other day* although **he is a vegetarian**.

Let’s consider (12), from the interrogative pattern. Whereas (13a) is a perfectly acceptable answer to it, (13b) is not.[6] This tends to show that with *although*, Arg$_2$ and R are not at-issue and that the interrogation only concerns the content of the matrix clause.

(12) Did he eat meat although he is a vegetarian?

[5] The examples in (11) would be more compelling if *although* was replaced with *despite the fact that.* However, the lack of appropriateness of *although* comes from subtle differences in semantics and usage which are unrelated to the problem at stake. It is for the sake of simplicity and homogeneity that we have chosen to stick to *although.*

[6] It would be possible to continue (12) with *He is not a vegetarian anymore*, but this is more of a remark than an answer: the dialogue could continue with *You haven’t answered my question.* Also note the use of *anymore*, which marks a revision.

(13) a. No, he refused.
b. #No, he is not a vegetarian anymore.

Yet, saying that the Arg_2 of a *Concession* is never at-issue would be taking shortcuts. It seems for example that (14a), from the AV pattern, can be felicitously followed by (14b) although it negates Arg_2. So in such a case, *he was sick* is under the semantic scope of *Sabine thinks*.

(14) a. Sabine thinks *Fred came to work* although **he was sick**.
b. But she is wrong, he had recovered several days ago.

Out of context, (14a) seems intuitively to imply that Fred was actually sick; this is a default reading. The utterance is ambiguous: the Arg_2 may or may not be under the semantic scope of the AV, the latter being the default interpretation.

Summary: Tab. 1 summaries these properties and those of two other SubConjs, *after* and *whereas*, which could not be discussed here due to lack of space. *although*, *whereas* and other SubConjs are ambiguous between *Contrast* and *Concession* (at least, [13]), but they have the same properties as long as they introduce a PAC.

It seems that conjunctions introducing a PAC ('peripheral conjunctions', PCs) all share the same behaviour; they allow mismatches for Arg_2, the speaker is always committed to the relation conveyed and in the (very probable) default reading the speaker is also committed to Arg_2. Conjunctions introducing a CAC ('central conjunctions', CCs) also share some properties; they do not allow any mismatch for Arg_2, the commitment of the speaker toward the relation conveyed is always subject to the modifiers used in the patterns. The status of Arg_2, however, depends on the conjunction.

R (type)	CONJ	Arg_2	$R(\text{Arg}_1, \text{Arg}_2)$	mismatch for Arg_2
Explanation (central)	because	−	−	−
Narration (central)	after	(+)	−	−
Concession (peripheral)	although	(+)	+	+
Contrast (peripheral)	whereas	(+)	+	+

Table 1. The Arg_2 and $R(\text{Arg}_1, \text{Arg}_2)$ columns show if the truth of these propositions are still implied by the use of CONJ in the studied patterns: '−' means 'no', '+' means 'yes' and '(+)' means 'yes in the default reading'; '+'/'−' in the last column indicate whether CONJ can or cannot be found with a mismatch concerning its Arg_2. Arg_1 is always subject to the operators used in the various patterns.

4 Our Proposition in STAG

We now turn to STAG and propose a basic model for AVs, CCs and PCs that reflects the properties observed in the previous sections. But before that, let us explain what this formalism is and how it works.

4.1 TAG and STAG

The Tree Adjoining Grammar formalism (TAG), on which is based STAG, was introduced in [10]. In TAG, words are represented as tree structures of two kinds. On the one hand are the *initial trees* (named with α), whose interior nodes are labelled with non-terminal symbols and whose leaves are either labelled with a terminal symbol, either labelled with a non-terminal symbol and marked with $\downarrow$. In the last case, the leaf is said to be a *substitution site*. On the other hand are the *auxiliary trees* (named with β), which are similar to initial trees except that they have (exactly) one leaf, the *foot node*, that is labelled with the same non-terminal as the root and is marked with $*$ instead of $\downarrow$.

These trees are meant to combine into sentences using two operations: *substitution* and *adjunction*. A set of such operations is called a *derivation tree* and the resulting tree is called a *derived tree*. A substitution consists in replacing a substitution site with an whole initial tree whose root must be labelled with the same symbol that the substitution site. An adjunction consists in inserting at some interior node an auxiliary tree whose root must also be labelled with the same symbol that the target node. Both operations are illustrated in the upper part of Fig. 1 with a syntactic grammar: a substitution and an adjunction are represented on the left side while the resulting tree appears on the right side.

The idea of STAG [15] is to pair two TAGs together to perform parallel operations. Thus, in STAG, a lexical entry is a pair of TAG trees with a set of links precising the coupling between the two. A link is a pair of nodes, one from each tree, here marked as $^{[1]}$, $^{[2]}$, etc. Only on a linked node can an adjunction or a substitution be performed. When a substitution (resp. an adjunction) occurs at a node, the parallel substitution (resp. adjunction) must also occur on the other node of the link. This principle is illustrated in Fig. 1 with the coupling of a syntactic grammar (top) and a semantic one (bottom).

Note that we allow multiple adjunctions on the same node – up to one for each link on that node. In such a case, the order of the adjunctions must be specified in order to describe the resulting derived tree pair. Otherwise, the derivation tree is *underspecified* and is used to represent all the derived trees corresponding to all possible orders.

4.2 Lexical Entries

AVs: To take into account the two *evidential* and *intentional* uses of AVs, we propose an initial TAG pair for these verbs in addition to the auxiliary one traditionally used (Fig. 2). Auxiliary trees for AVs are motivated by long distance extractions [10], where *John says* is equivalent to the adjunction of *according to John.* But this equivalence does not hold for intentional AVs in a discourse structure. Furthermore, adjunction is generally used to indicate semantic modifiers, whereas an intentional AV provides the main predicate that is argument of a DR and does not merely indicate attribution. That is why we think it makes sense to model them with initial trees rather than auxiliary trees.

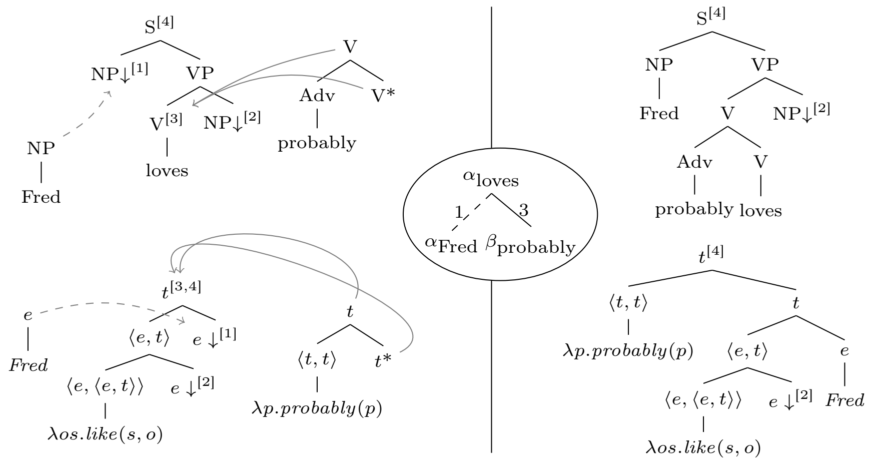


Figure 1. Substitution of α_{Fred} at link [1] in α_{loves} and adjunction of β_{probably} at link [3]; circled in the middle is the derivation tree; the derived trees are on the right.

In our model their semantics is also slightly different: evidential AVs use predicates that are 'erased' when in a DR. This is achieved by introducing rewriting rules of the form $Contrast(p, think(a, q)) \to Contrast(p, q)$. Conversely, unnatural mismatches can be avoided by discarding any analysis displaying an evidential AV predicate as argument of a central DR: $Explanation(p, think(a, q)) \to \bot$.

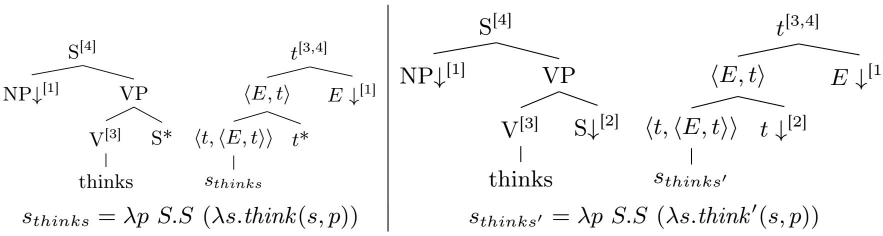


Figure 2. AVs: β_{thinks} (evidential) and α_{thinks} (intentional)

Subordinate conjuctions: Similarly, the difference in syntax and semantics between CACs and PACs can be explained with different structures for CCs and PCs as in Fig. 3.[7] The most significant aspect of these structures is that CCs are auxiliary trees whereas PCs are initial trees.

[7] The presence of the SBAR node for CCs is necessary because of the possibility of cleft sentences (*It is because A that B*), which shows that there exists such a constituent. No cleft sentences are observed with PCs.

Indeed, consider a CC adjoined to its matrix clause. If a semantic modifier (a negation, for instance) is also adjoined to this clause, depending on where this adjunction is done relatively to the adjunction of the CC, this modifier may or may not scope over the DR. With a PC, however, because the matrix clause is substituted into the connective, any modifier is necessarily dominated by the connective and thus only the local scope is possible.

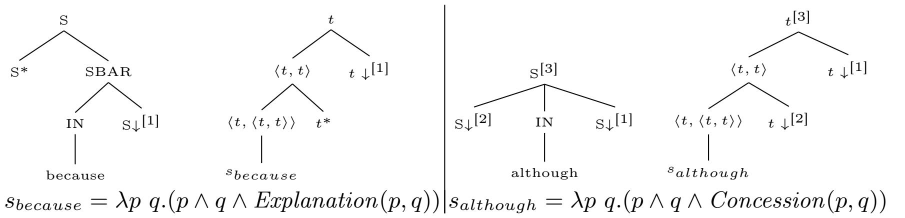


Figure 3. SubConjs: β_{because} (CC) and α_{although} (PC)

Sentence structures: [11] proposes that in sentence structures, verbal modifiers are adjoined in the semantic tree at a lower node than AVs. Doing so avoids (unnatural) interpretations of the former scoping over the latter. However, as seen in the previous section, while CCs are sentence modifiers like AVs they do present scope ambiguity when confronted with verbal modifiers such as negation. This is why, as illustrated in Fig. 4, we consider adding to sentence structures another adjunction site on the S-node (link [3]) whose semantic counterpart is at the same node as verbal modifiers' one. We can use features to restrict the other S-site (link [2]) to AVs and conversely to force them to adjoin there.

Fig. 4 also shows the derivation trees obtained when a negation (or any other verbal modifier) is present in the matrix clause of a SubConj. As expected, the negation can either have a local or a global scope with a CC while it always has local scope with a PC.

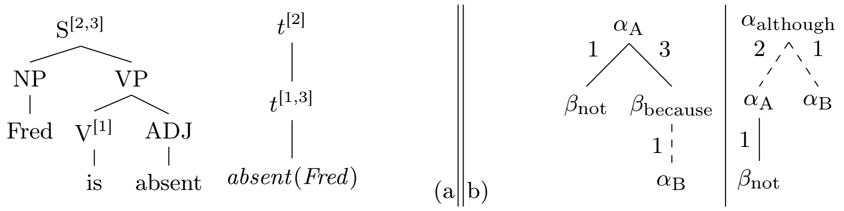


Figure 4. A typical sentence structure (a), accompanied by the derivation trees for *not A CONJ B* with a CC (b. left) or a PC (b. right). Because in α_{A} links [1] and [3] are at the same semantic node, the left tree is a scope neutral representation yielding one syntactic tree but two semantic ones depending on the order of the adjunctions.

5 Discussion and Perspectives

Tab. 2 shows the derivation trees for sentences of the form *A CONJ Sabine thinks B* as in (2). While the rewriting rules we have introduced discard the use of an evidential AV with a CC, there is an ambiguity for PC that can only be resolved using the particular semantics of the lexicalised relations.

Tab. 3 shows the derivation trees for sentences of the form *Sabine thinks A CONJ B* as in (14a). In this configuration, it is possible for the AV to scope over the relation. Its evidential or intentional status is then undetermined (it would depend on another discourse relation) and we have arbitrarily chosen to represent it with the traditional auxiliary structure in the 'External AV' column. As previously, the various ambiguities are natural ones and can only be resolved using world knowledge.

These analyses generate the expected derived trees, on both the syntactic and the semantic side (though we lack space to exhibit them here). However, our model should be refined: it does not yet account for the projection of peripheral DRs as observed in (14b) nor for the default projection of Arg_2 for some CCs such as *after* (not discussed here, but that some also see in *because* - see note 4).

Furthermore, discussing the interaction between AVs and SubConjs, we have said that evidential AVs could be replaced with attributing prepositional phrases such as *according to Sabine* (15a). The other way around is not true, because such expressions can be found within CACs (15b), where evidential are forbidden. In fact, it seems that in these cases the attribution not only scopes over Arg_2 but also over the relation itself. Attributing prepositional phrases therefore exhibit very interesting behaviours at the discourse level and will likely prove challenging to model considering their relatively free position in the sentence (15).

We also intend on extending this study to other connectives (especially adverbials, such as *instead* and *otherwise*) with the ultimate goal of building a parser capable of providing analyses coherent at the syntactic, semantic and discourse levels.

(15) a. *Fred could not come* even though, according to Sabine, **he was really looking forward to it**.
b. *Fred could not come* because, according to Sabine, **he was not in town**.
c. *Fred could not come* even though **he was**, according to Sabine, **really looking forward to it**.
d. *Fred could not come* even though **he was really looking forward to it**, according to Sabine.

Acknowledgements

I would like to particularly thank my supervisor Laurence Danlos for her invaluable advice, as well as my friends Rachel and Emily for their help with the English language.

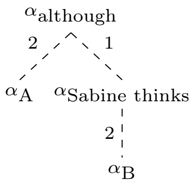
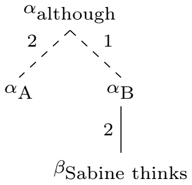
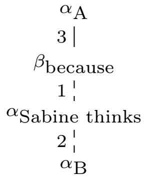
	Intentional AV	Evidential AV
PC:	 α_{although} 2 1 α_{A} $\alpha_{\text{Sabine thinks}}$ 2 α_{B}	 α_{although} 2 1 α_{A} α_{B} 2 $\beta_{\text{Sabine thinks}}$
CC:	 α_{A} 3 β_{because} 1 $\alpha_{\text{Sabine thinks}}$ 2 α_{B}	╳

Table 2. Derivation trees for sentences of the form *A CONJ Sabine thinks B*.

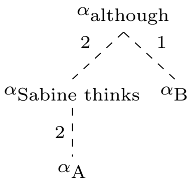
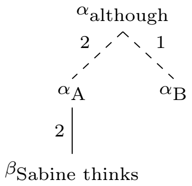
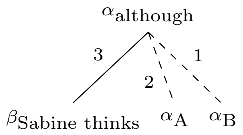
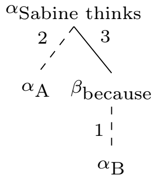
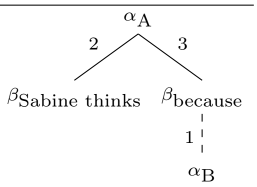
	Intentional AV	Evidential AV	External AV
PC:	 α_{although} 2 1 $\alpha_{\text{Sabine thinks}}$ α_{B} 2 α_{A}	 α_{although} 2 1 α_{A} α_{B} 2 $\beta_{\text{Sabine thinks}}$	 α_{although} 3 1 2 $\beta_{\text{Sabine thinks}}$ α_{A} α_{B}
CC:	 $\alpha_{\text{Sabine thinks}}$ 2 3 α_{A} β_{because} 1 α_{B}	╳	 α_{A} 2 3 $\beta_{\text{Sabine thinks}}$ β_{because} 1 α_{B}

Table 3. Derivation trees for sentences of the form *Sabine thinks A CONJ B*.

References

1. N. Asher, J. Hunter, P. Denis, and B. Reese. Evidentiality and intensionality: Two uses of reportative constructions in discourse. In *Workshop on Constraints in Discourse Structure*, Maynooth, Ireland, 2006.
2. G. Chierchia and S. McConnell-Ginet. *Meaning and Grammar: An Introduction to Semantics.* MIT Press, 1990.
3. L. Danlos. D-STAG: a Formalism for Discourse Analysis based on SDRT and using Synchronous TAG. In P. de Groote, editor, *Proceedings of FG'09.* INRIA, 2009.
4. L. Danlos. Connecteurs de discours adverbiaux: Problèmes à l'interface syntaxe-sémantique. *Linguisticae Investigationes*, 36(2):261–275, December 2013.
5. N. Dinesh, A. Lee, E. Miltsakaki, R. Prasad, A. Joshi, and B. Webber. Attribution and the (Non-)Alignment of Syntactic and Discourse Arguments of Connectives. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 29–36, Ann Arbor, Michigan, June 2005. ACL.
6. G. Frege. Sense and Reference. *The Philosophical Review*, 57(3):209–230, 1948.
7. H. Grice. Logic and conversation. In P. Cole and L. Jerry, editors, *Syntax and semantics 3: Speech acts*, pages 41–58. Academic Press, San Diego, CA, 1975.
8. L. Haegeman. The syntax of adverbial clauses and its consequences for topicalisation. In M. Coene, G. De Cuyper, and Y. D'Hulst, editors, *Current Studies in Comparative Romance Linguistics*, number 107 in APiL, pages 61–90. Antwerp University, 2004.
9. J. Hunter and L. Danlos. Because We Say So. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language*, CAtoCL, pages 1–9, Gothenburg, Sweden, April 2014. ACL.
10. A. Joshi. An introduction to tree adjoining grammars. *Mathematics of language*, 1:87–115, 1987.
11. R. Nesson and S. Shieber. Simpler TAG semantics through synchronization. In *Proceedings of FG 2006*, pages 129–142, Malaga, Spain, 2006.
12. C. Potts. Presupposition and Implicature. In S. Lappin and C. Fox, editors, *The Handbook of Contemporary Semantic Theory*, pages 168–202. Wiley-Blackwell, 2 edition, 2015.
13. R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. The Penn Discourse TreeBank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrackech, Morocco, June 2008.
14. B. Shanon. On the Two Kinds of Presuppositions in Natural Language. *Foundations of Language*, 14(2):247–249, 1976.
15. S. Shieber and Y. Schabes. Synchronous Tree-adjoining Grammars. In *Proceedings of the 13th Conference on Computational Linguistics - Volume 3*, COLING '90, pages 253–258, Stroudsburg, PA, USA, 1990. ACL.

How Reading Intercomprehension Works among Slavic Languages with Cyrillic Script

A Comparative Analysis of Common Slavic Vocabulary with the Focus on Orthographic Intelligibility

Irina Stenger

Saarland University, Collaborative Research Center (SFB) 1102:
Information Density and Linguistic Encoding
Project C4: INCOMSLAV
Mutual Intelligibility and Surprisal in Slavic Intercomprehension
ira.stenger@mx.uni-saarland.de
http://www.sfb1102.uni-saarland.de

Abstract. This article presents methods and results of a comparative analysis of Slavic languages using the Cyrillic alphabet with the purpose of investigating the orthographic intelligibility in reading intercomprehension. Although many studies have focused on linguistic as well as non-linguistic factors of mutual intelligibility among related languages, the understanding of the role of orthography in this process is quite limited. This study attempts to reveal the main mechanisms and the basic units of the orthographic code with focus on similarities between related languages, which leading to orthographic transparency and thus facilitating reading intercomprehension.

1 Introduction

As a special mode of language use, intercomprehension (or receptive multilingualism) is defined as "a form of communication in which each person uses his or her own language and understands that of the other" (Doyé 2005). The similarities in phonology, morphology, syntax and basic vocabulary among Slavic languages are striking (Carlton 1991). However, Townsend and Janda (1996:25) point out that "[m]ost Slavs speak of understanding each other without much difficulty, but this is usually exaggerated and applies mostly to a simple concrete level". In an intercomprehension scenario it is not exactly clear how the information of a message encoded in one language system is decoded by readers used to a different but genetically related language system. The focus of the present study lies on intercomprehension that refers to the reading competence. Various reading studies have evidenced that reading is a complex and structured multilevel process that involves not only knowledge of linguistic elements, but also the entire knowledge of the reader (Frost 2012; Lutjeharms 2002). While the degree of intelligibility of an unknown but closely related language depends on both linguistic and extra-linguistic factors (Gooskens 2013), the orthography represents a primary linguistic interface for any reading activity. The intelligibility of the script can therefore facilitate or complicate reading intercomprehension in case of related languages. A comparison of different orthographic systems involves various descriptive linguistic levels (i.e., phonetics/phonology, graphemics/graphotactics, morphology/morphosyntax, semantics) as well as historical, etymological and sociolinguistic factors (e.g., spelling reforms) (Sgall 2006). This study addresses the following questions: What are the main mechanisms and the basic units of orthographic code? How can orthographic similarities and differences be defined for a language pair? To what extent are the Slavic languages with the Cyrillic script orthographically intelligible?

2 Previous research

Sometimes readers are confronted with texts written in an unknown but closely related language and may be able to understand the text to some extent (cf. Heeringa et al. 2014). In order to understand a text, it is necessary to process the data contained in the text itself (bottom-up) and to link it with previous knowledge (top down). There is a permanent correlation between these two processes and the interaction takes place at various levels (Lutjeharms 2002). In our research project INCOMSLAV[1] we investigate reading intercomprehension among Slavic languages and approach the problem of their mutual intelligibility from an information-theoretic perspective in terms of surprisal, taking into consideration information en- and decoding at different linguistic levels. Our aim is to bring together the results from the analysis of parallel corpora and from a variety of experiments with native speakers of Slavic languages. Within the project, we carried out large-scale computational transformation experiments on parallel word sets (Pan-Slavic lists, internationalism lists, Swadesh lists)[2] with orthographic correspondences based on traditional approaches and comparative historical linguistics (Fischer et al. 2015). This allows us to investigate to what degree Czech and Polish (both West Slavic, using the Latin alphabet) and Bulgarian and Russian (South and East Slavic, both using the Cyrillic alphabet) are mutually intelligible at the orthographic level and to analyze the most frequent orthographic correspondences within the respective Slavic language pairs. Analyzing the results, we noticed a significantly higher rate of orthographically identical items between Bulgarian and Russian than between Czech and Polish (max.: 33.21% for CS-PL vs. 62.45% for BG-RU, both in the internationalism lists). Additionally, we measured the average orthographic distance within the Slavic language pairs by means of the Lenveshtein algorithm. Levenshtein distance is equal to the number of operations needed to transform one string of

[1] INCOMSLAV – Mutual Intelligibility and Surprisal in Slavic Intercomprehension. The project is part of the Collaborative Research Center (SFB) 1102 Information Density and Linguistic Encoding at Saarland University, launched in October 2014.

[2] http://www.eurocomslav.de/kurs/pwslav.htm;
http://www.eurocomslav.de/kurs/iwslav.htm;
http://en.wiktionary.org/wiki/Appendix:Swadesh_lists_for_Slavic_languages.

characters into another. There are three types of operations: insertions, deletions and substitutions. For example, in order to transform the Russian word "*рыба*" (*ryba*) into the Bulgarian "*риба*" (*riba*) "*fish*" the following step is necessary: replace *ы* with *и*. This operation costs one point and is the non-normalised Levenshtein distance. The normalised Levenshtein distance is obtained by dividing the non-normalised distance by the length of the longest alignment, which gives the minimum costs (cf. Beijering 2008). In our example we get 4 alignment slots and the normalised Levenshtein distance ist 1/4=0.25 or 25%. Using the Levenshtein algorithm, we found out that the average orthographic distance between Bulgarian and Russian is less than between Czech and Polish throughout all lists (cf. Stenger et al. forthcoming). At the next step we examined the lexical distance of the 100 most frequent nouns between four selected Slavic languages (Czech, Polish, Bulgarian, and Russian) with special attention to orthography.[3] Again the result was a smaller orthographic distance of respective cognate pairs by means of the Levenshtein algorithm between BG and RU than between CZ and PL with further asymmetries depending on the decoding direction for both pairs. Testing the mutual intelligibility between West (Czech, Slovak and Polish) and South Slavic languages (Croatian, Slovene and Bulgarian) among native speakers, Golubović and Gooskens (2015) found out that Czech, Slovak and Polish native speakers (West Slavic, all using the Latin alphabet), who indicated in the background questionnaire that they could read Cyrillic (might have learned another Slavic language written in Cyrillic, such as Russian), were more successful at translating isolated words from Bulgarian (South Slavic, using the Cyrillic alphabet) than they were with Croatian and Slovene words (South Slavic, both using the Latin alphabet). The Polish native speakers could translate even more words from Bulgarian than from Czech and Slovak. Based on the insights we gained from our previous research work and taking into consideration the results of other studies on receptive multilingualism focusing on orthography (cf. Heeringa et al. 2013, 2014; Möller, Zeevaert 2015; Vanhove 2015 etc.), this paper attempts to investigate the Cyrillic orthographic code more precisely and to determine the degree of orthographic intelligibility between related Slavic languages using the Cyrillic script.

3 Theoretical basis for the comparative method

3.1 Slavic languages and writing systems

From a diachronic perspective, there are three major scripts in Slavic area: Glagolitic, Cyrillic and Latin. "The Glagolitic alphabet was created by Constantine-Cyril in the 9th c. to render Slavonic [...], structurally influenced by the Greek alphabet. It was eventually replaced by Cyrillic (in Bulgaria) and by the

[3] Jágrová, K., Stenger, I., Marti, R., Avgustinova, T. (accepted Olinco 2016): Lexical and Orthographic Distances between Czech, Polish, Russian, and Bulgarian – a Comparative Analysis of the Most Frequent Nouns.

Latin alphabet (later in Croatia)." (Marti 2014:1497). "Like Glagolitic and unlike the Latin alphabet, Cyrillic was a script customized to the contemporaneous Slavic languages, with a highly efficient and systematic one-to-one correspondence between its graphemes and the Slavic set of phonemes." (Kučera 2009:72). "Cyrillic was used to write Church Slavonic and developed regional forms. A major formal change was introduced in Russia at the beginning of the 18th c. by the Petrine reforms (graždanskij šrift), bringing Cyrillic formally closer to the Latin script. The new script was eventually adapted by all Slavs using Cyrillic, adding in the process new letters and/or changing the letter-sound correlations." (Marti 2014:1497). Modern Slavic languages use two different alphabets: Latin and Cyrillic, which provide the Slavic orthographies with two rather different bases (Kučera 2009). This study focuses on East (Russian, Ukrainian, Belarusian) and South Slavic (Bulgarian, Macedonian, Serbian) languages, since they all use the Cyrillic alphabet. Historically Belarusian also used the Latin alphabet. Partly this holds true for Serbian as well.

3.2 Slavic orthographic principles as mechanisms of orthographic code

The main lines of sound system evolution in the Slavic languages are not always reflected in writing. On the one hand, the orthography could be adapted to sound changes in order to achieve harmony with the spoken languages; on the other hand, it could be based on the morphological principle (e.g. with morphemes being always written the same way despite differences in pronunciation) or historical/etymological principles (e.g. reflecting either the spelling of a language from which a word has been borrowed, or an older state of the same language) (Kučera 2009). Most, if not all, Slavic orthographies can be primarily described as phonemic (Kučera prefers the term "phonological"), i.e. based on the correspondences between graphemes and phonemes (Kučera 2009). However, we need to distinguish between phonemic and phonetic principles, the latter being based on correspondences between graphemes and sounds that are actually pronounced, like the spelling of the Russian prefixes ***раз-/рас-*** (***raz-/ras-***)[4] e.g., RU: ***развод***, ***распределение*** - [ra**z**][5]*вод*, [ra**s**]*пределение* (***razvod***, ***raspredelenie***) - in contrast to the spelling of the Bulgarian prefix ***раз-*** (***raz-***), e.g., ***развод***, ***разпределение*** - [ra**z**]*вод*, [ra**s**]*пределение* (***razvod***, ***razpredelenie*** "*divorce*", "*distribution*").

In different Slavic languages the development of structured orthographies followed different paths at different times, and the established writing systems can be considered as a result of both linguistic and sociolinguistic factors (Sgall 2006). A very important modification of the "civil alphabet" (graždanskij šrift) was effected by Vuk Karadžić in 1814/1818 for Serbian. The Serbian orthography

[4] Transliteration of Russian, Bulgarian, Macedonian, and Belarusian words is given according to DIN 1460.

[5] Phonetic transcription of Russian, Bulgarian, Macedonian, and Belarusian is given according to IPA.

adheres to the phonemic principle, with a strong tendency towards the phonetic principle (Kučera 2009; Marti 2014). The Russian orthography reform, which took place in 1918, eliminated a number of letters and changed spelling rules. Despite the fact that Russian orthography is based in general on the phonemic principle (Ivanova 1991), the morphological principle is considered to be important (e.g., the root in the following examples is always written the same way despite of differences in pronunciation: ***ход***, ***ход**óк*, ***ход**овóй* - **[xot]**, **[xʌd]***óк*,**[xəд]***овóй* (***chod***, ***chod**ók*, ***chod**ovój* derived from the morpheme ***chod-*** "*walk*") (Valgina 2002). The Bulgarian orthography (and this holds true for Ukrainian and the Belarusian as well) generally followed the Russian model with a number of changes in the alphabets (Kempgen 2009; Marti 2014). The Bulgarian orthography is defined as phonemic, although the morphological principle is also very important (Maslov 1981). The Ukrainian orthography is considered to be morphophonemic (Žovtobrjuch, Moldovan 2005) and the Belarusian orthography is based on two main principles: phonetic and morphological (Birillo et al. 2005). A phonetic representation of the vowels, but not of the consonants is characteristic for Belarusian (e.g., the unstressed *o* changes to *a*: *гарá - гóры* (*hará - hóry* "*mountain - mountains*"). Macedonian adapted the Serbian alphabet following the phonemic principle of Serbian Cyrillic (also with a strong tendency towards the phonetic principle) with some exceptions according to the morphological principle (Usikova 2003).

The Cyrillic orthographies based primarily on the phonemic (or phonological) principle make also use of other orthographic principles (e.g., phonetic, morphological, historical/etymological). All of them represent nowadays so-called mixed systems providing the respective languages with a number of general patterns. The hypothesis is that the adapted orthographic principles as mechanisms of orthographic code play a significant role in reading intercomprehension. In spite of differences, for example, on the phonetic/phonological level between the selected Slavic languages, the particular orthographic principles used in the writing systems may lead to formally identical orthographic representations of linguistic items. Within a language pair, the formal orthographic identity can be seen as orthographic similarity leading to script transparency and may increase the mutual intelligibility (e.g., differently pronounced linguistic items in Russian and Macedonian [vʌda] vs. [vɔda] are orthographically identical and therefore transparent: *вода - вода* (*voda* "*water*")).

3.3 Cyrillic alphabets as the basis of the orthographic code

An alphabet of a language can be described as a set of letters (graphemes) that is used to constitute a written text in the language (Sgall 2006). In other words, an alphabet represents a basis for the orthography and a letter (grapheme) can be defined as a basic unit of the orthographic code. The selected modern Slavic languages - Russian, Ukrainian, and Belarusian (East Slavic languages) and Bulgarian, Macedonian, and Serbian (South Slavic languages) - use the following Cyrillic alphabets[6]:

Russian:

А а, Б б, В в, Г г, Д д, Е е, Ё ё[7], Ж ж, З з, И и, Й й, К к, Л л, М м, Н н, О о, П п, Р р, С с, Т т, У у, Ф ф, Х х, Ц ц, Ч ч, Ш ш, Щ щ, Ъ ъ, Ы ы, Ь ь, Э э, Ю ю, Я я (33)

Ukrainian:

А а, Б б, В в, Г г, Ґ ґ, Д д, Е е, Є є, Ж ж, З з, И и, І і, Ї ї, Й й, К к, Л л, М м, Н н, О о, П п, Р р, С с, Т т, У у, Ф ф, Х х, Ц ц, Ч ч, Ш ш, Щ щ, Ь ь, Ю ю, Я я (33)

Belarusian:

А а, Б б, В в, Г г, Д д, Е е, Ё ё, Ж ж, З з, І і, Й й, К к, Л л, М м, Н н, О о, П п, Р р, С с, Т т, У у, Ў ў, Ф ф, Х х, Ц ц, Ч ч, Ш ш, Ы ы, Ь ь, Э э, Ю ю, Я я (32)

Bulgarian:

А а, Б б, В в, Г г, Д д, Е е, Ж ж, З з, И и, Й й, К к, Л л, М м, Н н, О о, П п, Р р, С с, Т т, У у, Ф ф, Х х, Ц ц, Ч ч, Ш ш, Щ щ, Ъ ъ, Ь ь, Ю ю, Я я (30)

Macedonian:

А а, Б б, В в, Г г, Д д, Ѓ ѓ, Е е, Ж ж, З з, Ѕ ѕ, И и, Ј ј, К к, Л л, Љ љ, М м, Н н, Њ њ, О о, П п, Р р, С с, Т т, Ќ ќ, У у, Ф ф, Х х, Ц ц, Ч ч, Џ џ, Ш ш (31)

Serbian:

А а, Б б, В в, Г г, Д д, Ђ ђ, Е е, Ж ж, З з, И и, Ј ј, К к, Л л, Љ љ, М м, Н н, Њ њ, О о, П п, Р р, С с, Т т, Ћ ћ, У у, Ф ф, Х х, Ц ц, Ч ч, Џ џ, Ш ш (30)

The use of digraphs as well as of diacritics is rare in the Cyrillic script. Some linguists add, for example, the digraphs *Дз дз* and *Дж дж* to the Belarusian alphabet (Birillo et al. 2005). The apostrophe in Belarusian and Ukrainian as an equivalent to the Russian letter *ъ* is not listed in the Belarusian and Ukrainian

[6] Montenegrin as a South Slavic language also uses the Cyrillic alphabet, but it will not be considered here. The different varieties of Rusyn using Cyrillic will be excluded, too.

[7] The letter *ë* is generally used in dictionaries and schoolbooks only.

alphabets. The number of letters in each alphabet is different and ranges from 30 to 33 symbols. Comparing, for example, Bulgarian and Russian alphabets (these Slavic languages belong to different sub-groups) we see that there are only slight differences in the alphabets. Three letters of the Russian alphabet do not occur in Bulgarian: *ы, э, ё*. The forms of the capital and small Bulgarian letters (graphic characters of the alphabet printed here) do not differ from their Russian counterparts (e.g., BG-RU: *А, а* : *А, а*). In an intercomprehension scenario all Bulgarian letters seem to be familiar to readers who know the Russian alphabet, but not vice versa. However, the nature as well as the use and pronunciation of a number of Bulgarian letters are not the same as in Russian. The next step is to see how the formally identical letters are used in a written text in order to reveal orthographic similarities in a language pair based on straight (one-to-one) matching of orthographic correspondences (e.g., RU-UK: *а:а*, *о:о*; MK-SR: *р:р*, *њ:њ*; BG-BE: *д:д*, *й:й* etc.).

4 Comparative method

Successful reading intercomprehension is very closely linked to the amount of common vocabulary among genetically related languages (Möller, Zeevaert 2015). However, very often these etymologically related cognates are not identical (e.g. orthographically). Between Slavic cognates, differences can be rather large, as a consequence of the development of Slavic languages from unity (Proto-Slavic or Common Slavic) to diversity (modern Slavic languages). Möller and Zeevaert (2015:314-315) point out that "minor or major differences must be bridged in order to map the input onto the corresponding L1 item in the mental lexicon. The degree of difficulty of this process - often referred to as transparency of the relationship between cognates - is probably one of the most important factors for success or failure of intercomprehension." Earlier research of orthographic intelligibility has mostly involved analyzing orthographic differences and measuring the orthographic distance between respective cognate pairs by means of the Levenshtein algorithm (cf. Heeringa et al. 2013, 2014; Möller, Zeevaert 2015; Vanhove 2015 etc.). The focus of this study lies on orthographic intelligibility in terms of orthographic similarity. The orthographic distance between orthographically identical items is zero by means of the Levenshtein algorithm. The comparative analysis intends to determine the degree of orthographic intelligibility between the selected Slavic languages with the practical aim to detect orthographic units that may facilitate the access to the text presented in a related language and be used as didactic material for this purpose. A set of one-to-one (matching) orthographic correspondences will be automatically extracted for 15 language pairs based on the part of orthographically identical items.

4.1 Data sources

To exclude the influence of other linguistic factors as far as possible the common (etymologically related) Slavic vocabulary (Carlton 1991) was adapted for the synchronic comparative analysis. The comparison of basic Slavic vocabulary (Carlton 1991) consists of 212 examples for 15 languages: Proto-Slavic, Old Church Slavic/Slavonic, Ukrainian, Russian, Belarusian, Bulgarian, Macedonian, Serbo-Croatian, Slovenian, Polish, Czech, Slovak, Upper Lusatian, Lower Lusatian, and Polabian. The parallel lists are divided into the following parts: (1) commonly used adjectives, (2) common animals and birds, (3) common plants, (4) commonly used verbs, (5) kinship terms, (6) nature, tools, housing, (7) terms referring to nourishment, (8) parts of the body (human and non-human). The Carlton's parallel vocabulary lists are based on the analysis of the Slavic inherited lexicon of Mel'nyčuk (1966). In order to exclude any typing mistakes both data sources were compared. Additionally, the etymological dictionaries of Žuravlev (1974-2011) and Vasmer (1973) were also used. While the original data sources have some empty slots for one or the other language, those parallel examples with and without empty slots are not included into the present analysis in order to compare consequently all 6 languages. The parallel vocabulary lists of this study include only 190 items, consisting mostly of nouns with a small amount of 23 adjectives and 27 verbs in each language.

4.2 Results of comparative analysis of common Slavic vocabulary

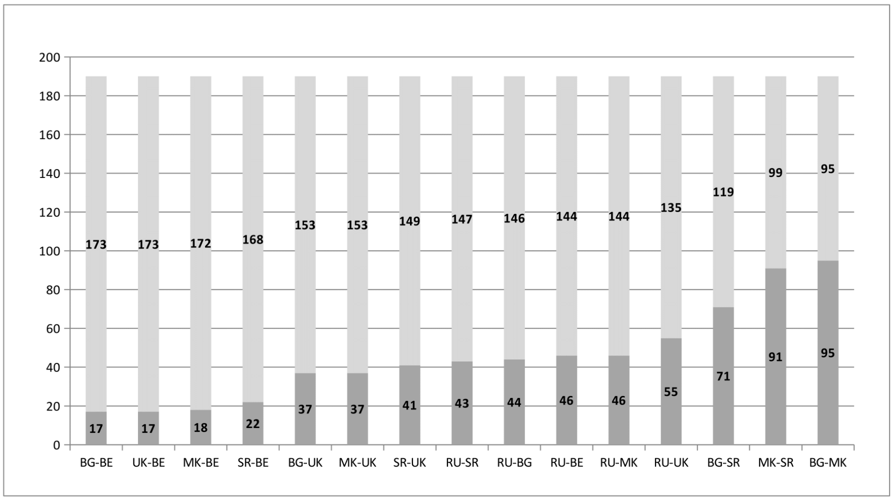
The automatic extraction of orthographically identical items and of one-to-one (matching) orthographic correspondences for 15 language pairs was done in the framework of the project INCOMSLAV based on Python 2.7.6. Diagram 1 shows the comparative analysis of common Slavic vocabulary for 15 language pairs. On the one hand, the BG-MK pair reveals the highest rate of orthographically identical items: 95. This result confirms the close relationship between the two languages. On the other hand, the BG-BE pair has the smallest rate of orthographically identical items: only 17. This can be explained, for example, by differences on the phonetic/phonological level that are reflected in writing according to the observed orthographic principles. For this language pair, the occurrence of different morphological features in the data set leads to certain discrepancies: e.g., (i) infinitive verb forms of Belarusian vs. 1st person forms of verbs in Bulgarian; (ii) different adjective forms: zero endings vs. different endings of adjectives in the masculine form for BG-BE; (iii) different ending of nouns etc.

Comparing, for example, Russian with other Slavic languages, we see that the rate of orthographically identical items is relatively constant and ranges from 43 (RU-SR) to 55 (RU-UK) items. This suggests a "bridge" position of Russian with regard to the other Slavic languages that use the Cyrillic alphabet.

An interesting finding is the smallest rate of orthographically identical items between Belarusian and other Slavic languages except for Russian. This positions Belarusian as the most distant language from the others in terms of orthographic intelligibility: the representation of unstressed vowels in the Belarusian plays a significant role.

Finally, another important result is the significantly higher rate of orthographically identical items within language pairs of South Slavic than within language pairs of East Slavic. Based on this, one could hypothesize that the degree of mutual intelligibility between South Slavic languages is higher than between East Slavic languages. However, this hypothesis reflects only the formal orthographic aspect of common Slavic vocabulary without taking into consideration the detailed semantic analysis of the compared etymologically related cognates.

Diagram 1
Comparative analysis of common Slavic vocabulary: orthographically identical and orthographically different items



■ - orthographically identical items (e.g., RU-BG: *брат-брат* (*brat-brat "brother"*))

□ - orthographically different items (e.g., RU-BG: *рыба-риба* (*ryba-riba "fish"*))

Additionally, a set of one-to-one (matching) orthographic correspondences was automatically extracted for 15 language pairs based on the part of orthographically identical items. The largest set of 27 one-to-one correspondences was obtained for RU-BE pair (e.g., *а:а*, *л:л*, *ё:ё*, etc.) and the smallest set of 19 matching correspondences was collected for MK-BE pair (e.g., *о:о*, *н:н*, *к:к*, etc.). This collection has to be completed with additional (mismatched) orthographic correspondences based on comparative historical linguistics (e.g., RU-BG: *ы:и*, *о:ъ*, etc.). Such a set of orthographic correspondences can be implemented to support cognate recognition.

5 Conclusion and outlook

According to the High Level Group on Multilingualism (HGLM 2007) which was established by the European Commission in 2006 (cf. Schüppert et al. 2015) the research topic of the present paper can be seen as a contribution to the investigation of crosslinguistic communication in Europe through receptive multilingualism. This article presents methods and results of a comparative analysis of Slavic languages using the Cyrillic alphabet with the purpose of investigating the orthographic intelligibility in reading intercomprehension. The study attempts to reveal the main mechanisms and the basic units of the Cyrillic orthographic code with a focus on orthographic similarity between related Slavic languages. The analysis in Section 3 shows that in spite of differences, for example, on the phonetic/phonological level between the selected Slavic languages, the particular orthographic principles used in the writing systems as mechanisms of the orthographic code may lead to formally identical orthographic representations of linguistic items. These are based on straight one-to-one (matching) orthographic correspondences consisting of formally identical letters as basic units of the orthographic code. The formal orthographic identity can be seen as orthographic similarity leading to script transparency and thus may facilitate the mutual intelligibility in reading comprehension between related languages.

The synchronic comparative analysis shows that the South Slavic language pairs have a significantly higher degree of orthographic similarity than the East Slavic language pairs. Orthographically identical items of the common Slavic vocabulary can be seen as a precondition that may facilitate the mutual intelligibility in a reading intercomprehension scenario. In addition, the results suggest that South Slavic languages are orthographically more intelligible to each other than East Slavic languages. Based on these findings one could hypothesize that a reader of one South Slavic language (using the Cyrillic script) may be more successful in understanding written texts in another related but unknown South Slavic language (also using the Cyrillic script) than an East Slavic reader being confronted with a written text in another unknown East Slavic language. This hypothesis has to be confirmed by further experiments taking into consideration the influence of other linguistic factors (e.g., morphology, syntax, semantics).

The research work described here is implemented in the framework of the INCOMSLAV project at Saarland University. The outcomes of this study will be the basis for computational transformation experiments with orthographic correspondences and will be tested in web-based reading intercomprehension experiments involving native Slavic speakers (e.g., free translation tasks, multiple choice – isolated words vs. words in context). The results will be used for building a feature-based language model mapping the encoding system of one language to another.

On a more applied note, reading intercomprehension can also be taught (cf. Zybatow 2002). The research results of the orthographic intelligibility between Slavic languages using the Cyrillic script can be used as didactic material in teaching and learning related Slavic languages with the purpose of showing the learners how much of the "completely new" language he/she already knows, without being aware of it.

Acknowledgement: I wish to thank Varvara Obolonchykova for her support in the automatic extraction of orthographically identical items and of one-to-one orthographic correspondences for 15 language pairs (based on Python 2.7.6). I am also grateful to Prof. Tania Avgustinova and Prof. Roland Marti for their explanations and helpful comments.

References

Beijering, K., Gooskens, C., Heeringa, W.: Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm. In: Linguistics in the Netherlands. John Benjamins Publishing Company, 13-24 (2008)

Birillo, N.V., Mackevič, Ju.F., Michnevič, E.E., Rogova, N.V.: Belorusskij jazyk. In: Moldovan, A.M. et al. (eds.) Jazyki mira. Slavjanskie jazyki. Academia, Moskva, 548-595 (2005)

Carlton, T.R.: Introduction to the Phonological History of the Slavic Languages. Slavica Publishers, Inc., Columbus, Ohio (1991)

Doyé, P.: Intercomprehension. Guide for the development of language education policies in Europe: from linguistic diversity to plurilingual education. Reference study. Strasbourg, DG IV, Council of Europe (2005)

Fischer, A., Jágrová, K., Stenger, I., Avgustinova T., Klakow, D., Marti R.: An Orthography Transformation Experiment with Czech-Polish and Bulgarian-Russian Parallel Word Sets. In: Sharp, B. et al. (eds.) Natural Language Processing and Cognitive Science 2015 Proceedings. Libreria Editrice Cafoscarina, Venezia, 115-126 (2015)

Frost, R.: Towards a universal model of reading. Behavioral and Brain Sciences. 35, (5). Cambridge University Press, 263-329 (2012)

Golubović, J., Gooskens, C.: Mutual intelligibility between West and South Slavic langauges. In: Russ Linguist 39. Springer, DOI 10.1007/s11185-015- 9150-9, 351-373 (2015)

Gooskens, C.: Experimental methods for measuring intelligibility of closely related language varieties. In: Bayley, R. et al. (eds.) Handbook of Sociolinguistics. Oxford University Press, 195-213 (2013)

Heeringa, W. Golubović, J., Gooskens, C., Schüppert, A., Swart, F., Voigt, S.: Lexical and orthographic distances between Germanic, Romance and Slavic languages and their relationship to geographic distance. In Gooskens C., van Bezoijen, R. (eds.) Phonetics in Europe: Perception and Production. Peter Lang, Frankfurt a.M., 99-137 (2013)

Heeringa, W., Swarte, F., Schüppert, A., Gooskens C.: Modeling intelligibility of written Germanic languages: do we need to distinguish between orthographic stems and affix variation? In: Journal of Germanic Linguistics 26 (4). Cambridge University Press, 361-394 (2014)

Ivanova, V.F.: Sovremennaja russkaja orfografija. "Vysšaja škola", Moskva (1991)

Jágrová, K., Stenger, I., Marti, R., Avgustinova, T. (accepted Olinco 2016): Lexical and Orthographic Distances between Czech, Polish, Russian, and Bulgarian – a Comparative Analysis of the Most Frequent Nouns.

Kempgen, S.: Phonetik, Phonologie, Orthographie, Flexionsmorphologie. In: Kempgen, S. et al. (eds.) The Slavica Languages. An International Handbook of their Structure, their History and their Investigation. Volume 1. Walter de Gruyter, Berlin/New York, 1-14 (2009)

Kučera, K.: The Orthographic Principles in the Slavic Languages: Phonetic/Phonological. In: Kempgen, S. et al. (eds.) The Slavic Languages. An International Handbook of their Structure, their History and their Investigation. Volume 1. Walter de Gruyter, Belin/New York, 70-76 (2009)

Lopatin, V.V.; Uluchanov, U.S.: Russkij jazyk. In: Moldovan, A.M. et al. (eds.) Jazyki mira. Slavjanskie jazyki. Acedemia, Moskva, 444-513 (2005)

Lutjeharms, M.: Lesestrategien und Interkomprehension in Sprachfamilien. In: Klein, H. G. et al. (eds.): EuroCom - Mehrsprachiges Europa durch Interkomprehension in Sprachfamilien. Shaker Verlag, Aachen, 119-136 (2002)

Marti, R.: Historische Graphematik des Slavischen: Glagoliitische und kyrillische Schrift. In: Gutschmidt, K. et al. (eds.) The Slavic Languages. An International Handbook of their Structure, their History and their Investigation. Volume 2. Walter de Gruyter, Berlin/New York, 1497-1514 (2014)

Maslov, J.S.: Grammatika bolgarskogo jazyka. "Vyssaja škola", Moskva (1981)

Mel'ničuk, O.S.: Vstup do porivnjal'no-istoryčnoho vyvčennja slov"jans'kich mov. Naukova dumka, Kiev (1966)

Möller, R.; Zeevaert, L.: Investigating word recognition in intercomprehension: Methods and findings. In: Linguistics 2015, 53(2). De Gruyter Mouton. Berlin/Munich/Boston, 313-352 (2015)

Sgall, P.: Towards a Theory of Phonemic Orthography. In: Sgall, P. (ed.) Language in its multifarious aspects. Karolinum Press, Charles University, 430-452 (2006)

Schüppert, A., Hilton, N.H., Gooskens, C.: Introduction: Communicating across linguistic borders. In: Linguistics 2015, 53 (2). De Gruyter Mouton, 211-217 (2015)

Stenger, I., Jágrová, K., Fischer, A., Avgustinova, T. (forthcoming): "Reading Polish with Czech Eyes" or "How Russian Can a Bulgarian Text Be?": Orthographic Differences as an Experimental Variable in Reading Comprehension. In: Kosta, P., Radeva-Bork, T. (eds.) (preliminary title) Current developments in Slavic Linguistics. Twenty years after (based on selected papers from FDSL 11 and a guest paper by N. Chomsky). Peter Lang

Townsend, C.E.; Janda, L.A.: Common and Comparative Slavic: Phonology and Inflection with special attention to Russian, Polish, Czech, Serbo-Croatian, Bulgarian. Slavica Publishers, Inc., Columbus, Ohio (1996)

Trofimkina, O.I.; Drakulič-Prijma, D.: Serbskij jazyk. Karo, Sankt-Peterburg (2011)

Usikova, R.P.: Grammatika makedonskogo literaturnogo jazyka. "Muravej", Moskva (2003)

Valgina, N.S., Rozental', D.E., Fomina, M.I.: Sovremennyj russkij jazyk. "Logos", Moskva (2002)

Vanhove, J.: The Early Learning of Interlingual Correspondences Rules in Receptive Multi-lingualism. In: International Journal of Biligualism http://ijb.sagepub.com/content/early/2015/03/05/1367006915573338 (2015)

Vasmer, M.: Etimologičeskij slovar' russkogo jazyka. "Progress", Moskva (1973)

Zybatow, L.: Slawistische Interkomprehensionsforschung und EuroComSlav. In: Klein, H. G. et al. (eds.) EuroCom - Mehrsprachiges Europa durch Interkomprehension in Sprachfamilien. Shaker Verlag, Aachen, 313-328 (2002)

Žovtobrjuch, M.A., Moldovan, A.M.: Ukrainskij jazyk. In: Moldovan, A.M. et al. (eds.) Jazyki mira. Slavjanskie jazyki. Academia, Moskva, 513-548 (2005)

Žuravlev, A. F. (ed.): Etimologičeskij slovar' slavjanskich jazykov. Vyp. 1-37. "Nauka", Moskva (1974-2011)

Combining syntactic patterns and Wikipedia's hierarchy of hyperlinks to extract meronymic relations

Debela Tesfaye, IT Doctoral Program, Addis Ababa University, Ethiopia

Abstract. We present here two methods for extraction of semantic relations between two words: (a) the first one relies solely on syntactic patterns. Unlike other syntactic pattern-based approaches, we combine patterns, determining their optimal combination to extract word pairs linked via a given semantic relation; (b) the second approach consists in combining syntactic patterns with the semantic information extracted from the Wikipedia *hyperlink hierarchy* of the constituent words. We have evaluated our approach with respect to SemEval 2007 (Task 4 test set) and WordNet. we get a F measure of 88.9% on a standard test-set, which is better than other reported approaches.

1 Introduction

The attempt to discover automatically semantic relations (SR) between words, or word pairs has attracted a number of researchers during the last decade and is justified by the number of applications needing this kind of information. Question Answering, Information Retrieval and Text Summarization are just some examples in case (Turney and Littman, 2005; Girju et al., 2005). SRs are one of the major components of ontologies and other formal knowledge representations. Hence, automatic extraction of SRs from textual data is important, all the more as it minimizes the labor-intensive phase of manual knowledge encoding, helping engineers to overcome the well-known knowledge acquisition bottleneck.

SRs extraction approaches can be categorized on the basis of the kind of information used. The method using only syntactic information relies on the extraction of word-level, phrase-level, or sentence-level syntactic information. This approach has been introduced by Hearst (1992) who showed that by using a small set of lexico-syntactic patterns (LSP) one could extract with high precision hypernym noun pairs. Similar methods have been used since then by (Auger and Barriere, 2008; Marshman and L'Homme, 2006). These authors reported results of high precision for some relations, for example hyponymy, noting poor recall which was low. Furthermore, the performance of this approach varies considerably depending on the type of relation considered (Ravichandran and Hovy, 2002, Girju et al., 2005.

An alternative to the syntactic approach is a method relying on the semantics features of a pair of words. Most researchers using this approach (Alicia, 2007; Hendrickx et.al, 2007) rely on information extracted from lexical resources like WN (Fellbaum, 1998). Alas, this method works only for languages having a resource equivalent to WN. Yet, even WN may pose a proble because of its low coverage across domains (tennis problem).

Hybrid approaches consist in the combination of syntactic patterns with the semantic features of the constituent words (Claudio, 2007; Girju et.al 2005). They tend to yield better results. However, their reliance on WN make them amenable to the same criticism as the ones just mentioned concerning WN. More recently Wikipedia based similarity measures have been proposed (Strube, et.al, 2006; Gabrilovich, and Markovitch, 2007). While this strategy produces excellent results, few attempts have been made to extract SRs (Nakayama et. al, 2007; Yulan et, al , 2007).

In this paper we propose two approaches to extract SRs: exploitation of the patterns learned from syntactic structures and the information of the constituent words from Wikipedia. Our contribution is twofold. First, we propose a novel technique for extracting optimal combination of syntactic patterns to extract SRs. Second, we propose an approach for disambiguating the syntactic patterns (say Meronymic patterns like NN1-has-NN2) constructing hyperlinks-hierarchy from Wikipedia pages.

2 Our approach

Previous works on syntactic structure are aimed at using unambiguous stand alone syntactic patterns to extract SRs. Even though these approaches achieved high precision, they are criticized for their low accuracy and the fact that their effectiveness greatly depends on the type of SRs to be extracted. One of the main challenges and research interest for syntactic pattern mining is how to disambiguate syntactic patterns for extracting SRs. In order to achieve this, we propose two approaches:

- Extracting optimal combination of *LSP*s to represent the relation at hand (section 2.1).
- Combining *LSP*s with the semantic features of the constituent words extracted from Wikipedia *hyperlinks-hierarchy* (section 2.2).

2.1 Combination of syntactic patterns for relation extraction (CoSP-FRe)

The use of individual syntactic patterns for the extraction of word pairs linked via a given SR produced poor results. One reason for this lies in the fact that the vast majority of word pairs are linked via polysemous syntactic patterns (Girju et.al , 2005). Hence, such patterns are not used alone, as they are ambiguous. At the same time they cannot be ignored as they have the potential to provide good clues concerning certain SRs. This being so, we suggest to assign weights to different *LSP*s according to their relevance for a specific SR, and to combine such weighted patterns for extracting word pairs linked via the relation at hand.

To determine the optimal combination of LSPs likely to extract SRs, we have harvested all syntactic patterns encoding the relation at hand. We assigned weights to the patterns according to their relevance for the given SRs, and finally we filtered the best combination of LSPs. We have extracted dependency grammar based LSPs encoding the word pairs linked via SR.

Determine the optimal combination of LSPs encoding some SR

To determine the optimal combination of LSPs, we used the discrimination value (dv) for each pattern. The dv, is closely related to tf-idf. It is a numerical value signaling how relevant a given LSP is with respect to a given SR. We applied the following steps in order to identify the dv and to determine the optimal combination of the LSPs:

Step 1: For each extracted LSP, we extracted more connected word pairs from a large corpus and built then word pairs in a LSPs matrix (Matrix 1). Next, we labeled the extracted word pairs with the SR type at hand and built a matrix of word pairs being linked by a specific SR type (Matrix 2). To this end, we automatically labeled the word pairs based on the type of SR as presented in WN (see algorithm 1). Using the information from Matrix 1 and 2, we built a matrix of SRs to LSPs (Matrix 3). The rows of the matrix are composed of the type of SRs. The columns represent the encoding LSPs. The cells are populated by the number of word pairs linked by the patterns encoding the SR.

The *dv* of *LSP* for a given SR is given by the following formula:

$$dv = \frac{fpr}{fp} * \log\left(\frac{tnr}{tre}\right) \qquad (1)$$

Where fp represents the total number of word pairs connected by the LSP (from Matrix 1). fpr represents the number of word pairs connected by the given SR (from Matrix 2), while tnr and tre represent respectively the total number of SRs (Matrix 3) and the total number of SRs encoded by the pattern (from Matrix 3). Frequent patterns unable to discriminate among the SRs are assigned a low dv and removed.

Step 2: Identify the optimal combination of LSP to represent the relation of interest. To achieve this, the matrix of combination of syntactic patterns by SRs is formed (Matrix 4) from matrix 3. The LSP are combined until no other combination is possible anymore. We have then calculated the discrimination value for the combined (grouped) syntactic patterns (dv-g). The dv-g is calculated for each combination of pattern corresponding to each SR. We selected the combination of patterns with maximum dv-g for each SR as described in algorithm 2. The dv-g for the combined patterns corresponding to a given SR is given by the formula:

$$dv - g = \frac{fpr - g}{fp - g} * \log\left(\frac{tnr}{tre - g}\right) \qquad (2)$$

fp-g expresses the total number of word pairs connected by the group of patterns. It is determined by the intersection of word pairs connected by the combined syntactic patterns (from Matrix 4)., fpr-g represents the number of word pairs connected by the combined patterns in the given SR. It is determined by the intersection of positive word pairs connected by the combined LSP for the given SR (from Matrix 4). Finally, tnr and tre represent respectively the total number of SRs (from Matrix 4) and the total number of SRs encoded by the combination of the LSP.

2.2 Wikipedia hyperlink hierarchies for SR extraction (WHH-Fre): Case of meronymy extraction

In this approach, we used the *hyperlink-hierarchies* constructed from selected sentences of Wikipedia pages of the respective word pairs to disambiguate syntactic patterns encoding them. The basic motivations behind this approach are as follows:

(a) Words linked to the Wikipedia page title (*wpt*) via *LSP* encoding SR are more reliable than word pairs linked in arbitrary sentences.

(b) SRs encoded by a given word pair can also be encoded by their respective higher/lower order conceptual terms. For instance:

(1). germ is an embryo of seed

(2). grain is a seed

would yield relations like *hyponymy (germ, embryo), hyponymy (grain, seed), meronymy (embryo, seed) and meronymy (germ, grain).* The *meronyms (germ, grain)* are inferred indirectly via the relation of their higher order terms *embryo* and *seed.*

The candidate meronymic word pairs extracted using meronymic LSPs are further refined using the patterns learned from their conceptual hierarchies constructed from two kinds of semantic links: (i) *hypernymic-link*, (ii) *meronymic-link*. To this end, we extracted hyperlinks connected to the Wikipedia pages of the respective meronymic candidates via *hypernymic* and *meronymic LSP*. We constructed the hyperlink hierarchies by analyzing only important sentences (1 and 2 below) from the Wikipedia pages of the pair of terms: (1) definition sentences and (2) sentences linking *hyperlinks* to the Wikipedia page title using meronymic patterns. Since the meronymic patterns vary according to their underlying nature, the patterns used to extract hyperlinks for constructing the hierarchies were learned taking the nature of the meronymic relations in to account (section 2.1). The definitions sentences are used to extract *hypernymic-hyperlink*[1] and the sentences linking hyperlinks to the Wikipedia page title using meronymic *LSP*s are used to extract *meronymic-hyperlink*[2].

Using the hierarchy constructed for the candidate word pairs, this approach determines whether the pairs are meronyms or not based on the following assumptions:

(a) The hyperlink hierarchies of *hierarchical meronymys* constructed form their respective *hypernymic-hyperlink* converges (have a common ancestor) along the

[1] hypernymic-hyperlink is a word defining a term using its higher order concept and providing hyperlink to other Wikipedia pages for further reading

[2] Meronymic-hyperlink is a word describing a term using its whole concept and providing a hyperlink to other Wikipedia pages for further reading

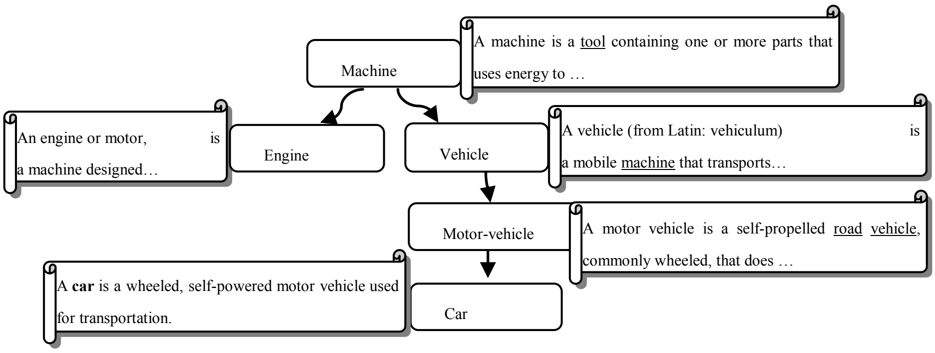
path in the hierarchy. Figure 1 shows the component-Integral meronyms *'car engine'* sharing the parent '*machine'* in their *hyperlink-hierarchy* constructed from their respective Wikipedia pages definitions.

(b) The hyperlink hierarchies of both *hierarchical and non-hierarchical-meronyms* constructed form their respective *meronymic-hyperlinks* and/or *hypernymic-hyperlink* converge along the path in the hierarchy.

Extracting the hyperlinks

In order to extract the *hyperlinks,* we took the following steps:

Step 1: For sample meronymic pairs we identified the respective Wikipedia pages. To achieve this, we aligned the word pairs with the *wpt* using simple word overlap. The word pairs were selected based on standard meronymic taxonomy (section 3.1.1). We first cleaned Wikipedia articles to remove unnecessary information such as HTML tags and special Wiki commands. Next, we extracted Wikipedia definitions and sentences linking *wpt* with hyperlinks using meronymic patterns.



A machine is a tool containing one or more parts that uses energy to …

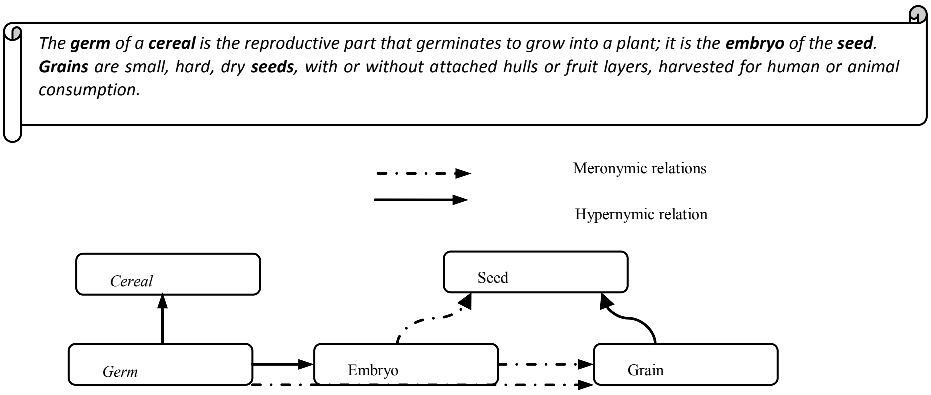
An engine or motor, is a machine designed…

A vehicle (from Latin: vehiculum) is a mobile machine that transports…

A motor vehicle is a self-propelled road vehicle, commonly wheeled, that does …

A **car** is a wheeled, self-powered motor vehicle used for transportation.

***Figure 1:** Wikipedia definitions and the resulting hypernymic-hyperlink hierarchies for the meronyms 'car engine'*



*The **germ** of a **cereal** is the reproductive part that germinates to grow into a plant; it is the **embryo** of the **seed**. **Grains** are small, hard, dry **seeds,** with or without attached hulls or fruit layers, harvested for human or animal consumption.*

***Figure 2:** Wikipedia definitions and the resulting hyponymic and meronymic hyperlink-hierarchies for the meronyms 'grain germ'*

Step 2: Anotations. We annotated both kinds of sentences relying on two kinds of information: *Wikipedia page title* and the *hyperlinks.* The *hyperlink* either links the term to its *meronyms* or *hypernyms*. The following Wikipedia sentences are annotated with their *Wikipedia page title* and the *hyperlinks* linking the term to its *hypernym* or *meronym* using *'wt', 'hl', 'ml'*respectively.

(3). *A car/wpt is a wheeled, self-powered motor vehicle/hl used for transportation.*

(4). *The germ/wpt of a cereal is the reproductive part that germinates to grow into a plant; it is the embryo/hl of the seed/ml.*

Step 3: Parsing. The annotated Wikipedia sentences were parsed to identify their dependency structure. The dependency grammar allowed us to learn the dependency based *LSP*s linking the *Wikipedia page title* with the *hyperlinks*.

Step 4: Extract dependency based syntactic patterns linking the *Wikipedia page title* with the *hyperlinks.* We assigned *dv (section 2.1)* for the patterns and considered then the top frequent patterns. The *hyperlinks* broadly fall in either of two categories: (a) *hypernymic-hyperlink.* They are extracted by the patterns linking the tuple (*hyperlink, Wikipedia page title*), for instance *is-a* (*hyperlink, Wikipedia page title)* (b) *meronymic-hyperlinks.* They are extracted using syntactic patterns linking the tuple (*hyperlink* , *Wikipedia page title*) for instance *made-from* (*hyperlink, Wikipedia page title)*.

Constructing the hierarchy

For a given pair of terms, we identified the respective Wikipedia pages aligning the pairs with the *wpt* using simple word overlap to extract their associated initial *hypernymic* and *meronymic hyperlinks* (hl_i) using the patterns learned in step 2.2.1. We further identified the respective Wikipedia pages for the *hypernymic* and *meronymic-hyperlink* from the previous step (hl_i) and extracted the associated *hypernymic* and *meronymic hyperlinks* (hl_{i+1}). We then connected (hl_i) with (hl_{i+1}) to form a hierarchy (*hypernyms* are connected to each other and to *meronyms*; *meronyms* are also connected to each other and to *hypernyms*). The hyperlinks are extracted until the hierarchies converge, or until the *hypernymic-hierarchy* reaches seven layers. Note that most word pairs converge before that. The precedence of the *hyperlinks* in the *hierarchy* is based on the order of appearance, the recent ones being the parent in the hierarchy. We performed stemming and lower cased the *hyperlinks* before extracting the respective sentences encoding them.

Decide whether two words are meronyms or not

Search the *hyperlink (hypernymic or meronymic-hyperlink)* of one of the words in the *hierarchy* of the other, and if so, consider the word pairs as meronyms. Figure 2 shows the meronymic word pairs '*germ grain*' converging at '*seed'* in their *hierarchy* built from their respective Wikipedia pages. Algorithm 3 shows WHH-FRe for the extraction of meronymy.

3 Experiment

To show the validity of our line of reasoning we carried out three experiments:

I. Extract the optimal combination of *LSPs* encoding meronymic relation.
II. Evaluate *CoSP-FRe* for meronymy extraction.
III. Evaluate *WHH-Fre* for extracting meronymy.

3.1 Extract the optimal combination of *LSP*s encoding meronymy

Training data set

Two sets of data are required: (a) the initial seed meronymic word pairs used to train our system (b) the corpus from which the syntactic patterns were selected. As mentioned in section 2, to select the representative list of meronymic pairs, we used a standard taxonomy. Indeed, several scholars have proposed taxonomies of meronyms (Winston et al., 1987; Pribbenow, 1995; Gerstl & Pribbenow, 1995; Vieu & Aurnague, 2007; Keet & Artale, 2008). We followed Winston's classical proposal:

1.	*component – integral-object(cio)*	*handle– cup*
2.	*member – collection(mc)*	*tree – forest*
3.	*portion – mass(pm)*	*grain – salt*
4.	*stuff – object(so)*	*steel – bike*
5.	*feature–activity(fa)*	*paying–shopping*
6.	*place-area(pa)*	*oasis–desert*

We used the part-whole training set of the SemEval-2007 task 4 (Girju et al. 2007) since it is organized following Winston et al.'s, (1987) meronymic taxonomy. The set contains 140 examples, of which 75 are negative and 65 positive. The set is POS-tagged and annotated with WN senses; We have removed these annotations.

We used the Wikipedia dump of 2013 for extracting the syntactic patterns encoded by the seed meronymics.

Experimental setup

The goal is, to determine the optimal combination of patterns encoding meronyms under each category by using the data set compiled in section 3.1.1. In order to achieve this goal we identified syntactic patterns encoding meronymy following the procedures described in section 2.1. Since the majority of the patterns are rare we considered only patterns with a frequency of 100 and above.

For the individual syntactic patterns extracted above, we have identified the *dv*s associated with the meronymic relation using *formula* 1 followed by the *dv-g*s for every combination of patterns using *formula 2*. The combined patterns are sorted based on their discrimination. Finally we selected the patterns with the highest *dv* as representatives of the respective meronymic types.

Sno	Pattern	Dv
1	NN_1 make of NN_2+ NN_2 to make NN_1 + NN_2 used NN_1 + NN_1 NN_2	83.6
2	NN_1 make from NN_2+ NN_2 to make NN_1 + NN_2 used NN_1 + NN_1 NN_2	81

Table 1. *Part of the optimal **c**ombination of patterns for so relations*

3.2 Evaluation

We have evaluated the performances of the two approaches (CoSP-FRe and WHH-FRe) by extracting the meronymic word pairs. The goal is to evaluate the degree of correspondence of the word pairs extracted as opposed to those by human annotators.

Test data set
We used two data sets: (a) the part-whole test set of the SemEval-2007 task 4 (Girju et al. 2007) which has 72 examples (26 positive and 46 negative) and WN's meronymic word pairs.

Comparison with other systems
We have compared our work against three approaches that achieved the best performance on SemEval-2007 task 4, and two other approaches. We categorized the approaches as WN based: CMU-AT (Alicia, 2007) & ILK (Hendrickx et.al, 2007), syntactic and hybrid approaches: FBK-IRST (Claudio, 2007) & Girjus et.al (2005). We used the individual *LSP*s (ILSP) extracted in Sections 2.1 and the *LSP*s extracted by Girju, et.al (2005) as syntactic approach. We used the LSPs to extract word pairs from the test set and compared it against the baseline.

Results
We computed precision, recall and F-measures as the performance metric. Precision is defined as the ratio of the number of correct meronyms extracted divided by the total number of word pairs extracted. Recall is defined by the ratio dividing the number of correct meronyms extracted by the total number of meronyms in the test set. Table 5 shows the precision, recall and F-measure of our algorithms and the other related state-of-the-art works. WHH-Fre proves to be highly reliable in extracting meronymy, achieving very competitive results.

	Data set		
	SemEval 2007		
Approaches	**P**	**R**	**F**
CoSP-FRe	76%	88%	81.5%
WHH-FRe	88%	90%	88.9%
ILSP	41.6%	87%	56.2%
CMU-AT	57.7%	45.5%	50.8%
FBK-IRST	65.5%	73.1%	69.1%
ILK	48.4 %	57.7%	52.6%

Table 3*: Recall (r), Precision (p) and F-Measure (f) of our approach and related works on SemEval test set*

We have also extracted meronymic word pairs from random Wikipedia pages of 100 articles. Out of the retrieved meronymic word pairs 85% are encoded in WN.

Discussions

We have discussed the results for both approaches in the following sections:

CoSP-FRe

The precision of CoSP-FRe is improved over syntactic approach as the ambiguity of the individual LSP's is reduced when patterns are combined. Recall is improved as a result of using ambiguous LSPs for extracting word pairs. This contrasts with all the other syntactic approaches which relied only on unambiguous LSPs. In our approach, ambiguous LSPs are also used in combination with other LSPs. Hence the coverage is significantly improved.

WHH-FRe

WHH-FRe outperforms significantly previous approaches both with respect to recall and precision as it combines two important features. First LSPs are used to extract lists of candidate pairs. Second semantic features of the constituent words extracted from Wikipedia hyperlink-hierarchy is used to further refine. Precision is improved for several reasons: relations encoding LSPs which link hyperlinks and WPT are more reliable than word pairs connected via arbitrary sentences. The features learned from the Wikipedia hyperlink-hierarchy further cleaned the word pairs extracted by LSPs. Recall is also improved since word pairs indirectly linked via their respective higher/lower order hierarchy were also extracted.

Based on the results of WHH-FRe several kinds of hierarchies were formed. Some of the hierarchies are made of hypernymic links or meronymic links and others are made from the combination of both links.

3.3 Related works

Syntactic approaches

The work of (Turney, 2005, 2006; Turney and Littman, 2005; Chklovski and Pantel, 2004) is closely related to our work (CoSP-Fre) as it also relies on the use of the distribution of syntactic patterns. However, their goals, algorithms and tasks are different. The work of (Turney, 2005, 2006; and Turney and Littma, 2005) is aimed at measuring relational similarity and is applied to the classification of word pairs (ex. *quart: volume* vs *mile: distance*) while we are aimed at extracting SRs.

Hybrid approaches

The work of Girju et.al (2005) is more related to our *WHH-FRe* in that they combined *LSPs* with the semantic analysis of the constituent words to disambiguate the *LSPs*. They used *WN* to get the semantics of the constituent words. Alicia (2007) converts word pairs of the positive examples into a semantic graph mapping the pairs to the *WN* hypernym hierarchy. Claudio (2007) combines information from syntactic processing and semantic information of the constituent words from *WN*. Wikipedia-based approaches mainly focused on the identification of similarity (Nakayama et. al, 2007; Yulan et, al , 2007). Also, there is hardly any recent work concerning the extraction of meronyms. Many researchers are working on the identification of semantic similarity achieving excellent result by using standard datasets (Camacho-Collados, Taher and Navigli, 2015; Taher and Navigli , 2015). Yet, most of this work dates back to 2010 and before.

4 Conclusions

We presented here two novel approaches for extracting SRs: CoSP-FRe and WHH-FRe. The strength of CoSP-FRe is its capacity to determine an optimal combination of *LSP*s in order to extract SRs. The approach yielded high precision and recall compared to other syntactic approaches. WHH-FRe matches the state of the art performance both with respect to recall and precision as it combines two important features: *LSP*s for extracting a list of candidate pairs and the use of semantic features of the constituent words extracted from Wikipedia *hyperlink-hierarchy*. Precision is improved for the following reasons: relation encoding *LSP*s linking hyperlink and Wikipedia page title are more reliable than word pairs connected via arbitrary sentences. Lexical semantic features learned from the Wikipedia *hyperlink-hierarchy* further cleaned the word pairs extracted by *LSP*s. Recall is also improved since word pairs indirectly linked via their respective higher/lower order hierarchy were also extracted. We plan to extend WHH-FRe to allow it to extract other SRs types than the ones dealt with in this paper, and in order to build a rich semantic network covering a large number of concepts.

5 References

Alain A. and Caroline B. (2008). Pattern-based approaches to semantic relation extraction: A state-of-the-art. Terminology Journal, 14(1):1–19

Alicia T. and Scott E. Fahlman (2007). CMU-AT: Semantic Distance and Background Knowledge for Identifying Semantic Relations, Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 121–124, Prague.

Cimiano, P., A. Pivk, L. Schmidt-Thieme and S. Staab. 2005. “Learning taxonomic relations from heterogeneous sources of evidence.” In Paul Buitelaar, Philipp Cimiano and Bernardo Magnini, Ontology Learning from Text: Methods, evaluation and applications. 55-73. Amsterdam: IOS Press.

Chklovski, T., and Pantel, P. (2004). VerbOcean: Mining the Web for fine-grained semantic verb relations. In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-04). pp. 33-40. Barcelona, Spain.

Claudio G., Alberto L., Daniele P. and Lorenza R. (2007). FBK-IRST: Kernel Methods for Semantic Relation Extraction, Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 121–124, Prague.

Fellbaum, C. editor. (1998). *WordNet: An electronic lexical database and some of its applications*. MIT Press.

Gabrilovich, E., Markovitch, S. (2007). Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: International Joint Conference on Artificial Intelligence, pp. 12-20.

Gerstl, P. & Pribbenow, S. (1995). *Midwinters, end games, and body parts: a classification of part-whole relations*. International Journal of Human-Computer Studies, 43, 865-889.

Girju R., Moldovan D., Tatu, M. & Antohe, D. (2005). *Automatic discovery of Part–Whole relations*. ACM 32(1)

Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., & Yuret, D. (2007). Semeval- 2007 task 04: Classification of semantic relations between nominals. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval 2007), pp. 13–18, Prague, Czech Republic.

Hearst, M. (1998). WordNet: An electronic lexical database and some of its applications. In Fellbaum, C., editor, Automated Discovery of WordNet Relations. MIT Press.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In Proceedings of the 14th International Conference on Computational Linguistics, pages 539–545.

Hendrickx I., Morante R., Sporleder C., Antal v. d. Bosch (2002). ILK: Machine learning of semantic relations with shallow features and almost no data. Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 121–124, Prague, June 2007

Jose Camacho-Collados, Mohammad Taher Pilehvar ´ and Roberto Navigli (205). NASARI: a Novel Approach to a Semantically-Aware Representation of Items, Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, pp 567–577, Denver, Colorado,USA.

Keet, C.M. and Artale, A. (2008). Representing and Reasoning over a Taxonomy of Part-Whole Relations. Applied Ontology, 2008, 3(1-2): 91-110

Malaisé, V., P. Zweigenbaum and B. Bachimont. 2005. Mining defining contexts to help structuring differential ontologies." Terminology 11(1): 21-53.

Marneffe M., MacCartney B. and Christopher D. Manning. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. In LREC 2006.

Marshman, E. and M.-C. L'Homme. 2006.Disambiguation of lexical markers of cause and effect" In Picht, H. (ed.). Modern Approaches to Terminological Theories and Applications. Proceedings of the 15th European Symposium on Language for Special Purposes, LSP 2005. 261-285. Bern: Peter Lang.

Moldovan D., Badulescu A., Tatu M., Antohe D., and Girju R. (2004). Models for the semantic classification of noun phrases. In Proc. of the HLT-NAACL 2004 Workshop on Computational Lexical Semantics, pages 60– 67, Boston, USA.

Morin, E. (1999). Automatic acquisition of semantic relations between terms from technical corpora. Proceedings of the 5th International Congress Terminology and Knowledge Extraction (TKE-99). 268-278. Vienna: TermNet.

Nakayama, K., Hara, T., and Nishio S. (2007). Wikipedia Mining for an Association Web Thesaurus Construction. In: Web Information Systems Engineering – WISE, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 322-334

Nakayama K., Hara T. and Nishio S. (2008). Wikipedia Link Structure and Text Mining for Semantic Relation Extraction. SemSearch 2008, CEUR Workshop Proceedings, ISSN 1613-0073, online at CEUR-WS.org/Vol-334/

Peter D. Turney and Michael L. Littman. (2005). Corpusbased learning of analogies and semantic rela-tions. Machine Learning, in press.

Peter D. Turney and Michael L. Littman. (2005). Corpus based learning of analogies and semantic relations. Machine Learning, 60(1–3):251–278

Peter D. Turney. (2006). Expressing implicit semantic relations without supervision. In Proceedings of ACL-2006.

Pribbenow, S. (1995). *Parts and Wholes and their Relations*. Habel, C. & Rickheit, G. (eds.): Mental Models in Discourse Processing and Problem Solving. John Benjamins Publishing Company, Amsterdam

Ravichandran, D. and E.H. Hovy. (2002). "Learning surface text patterns for a question answering system." In Proceedings of ACL-2002. 41-47. Philadelphia, Pensylvania.

Strube, M., and Ponzetto, S.P. (2006). WikiRelate! Computing semantic relatedness using Wikipedia. In: Proceedings of the National Conference on Artificial Intelligence, pp. 1419- 1429

Taher M. and Navigli R. (2015). *Align, Disambiguate and Walk: A Unified Approach for Measuring Semantic Similarity.* Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, , Sofia, Bulgaria, 1341–1351,

Vieu, L. & Aurnague, M. (2007). *Part-of relations, functionality and dependence*. In Aurnague, M., Hickmann, M. and Vieu, L. (eds.). The Categorization of Spatial Entities in Language and Cognition, pp. 307–336. J. Benjamins, Amsterdam

Winston, M., Chaffin, R. & Hermann, D. (1987). *Taxonomy of part-whole relations*. Cognitive Science, 11(4), 417–444.

Yan Y., Matsuo Y. , Ishizuka M. (2009). An Integrated Approach for Relation Extraction from Wikipedia Texts . CAW2.0 2009, Madrid, Spain.

Zesch, T., Müller, C., Gurevych, I. (2008). Extracting lexical semantic knowledge from wikipedia and wiktionary. In: Proceedings of the LREC, pp. 1646–1652.

(Co-)datatypes in HOL

Andrea Condoluci

Institut für Computersprachen, Technische Universität Wien
Favoritenstraße 9, 1040 Wien, Austria
andrea@logic.at

Abstract. We present a framework for defining freely generated inductive and co-inductive datatypes in the HOL proof assistant. The user can utilize previously defined types in new type specifications, as long as the functions 'map', 'index', and 'all' accompain these types. The three functions are already known in functional programming, share a similar format and are in continuation-passing style. Our method follows the traditional Melham–Gunter approach by representing types in trees, and it is going to be implemented in a new package for HOL that will allow mutual, nested, and mixed type definitions.

Keywords: higher-order logic, co-induction, datatypes

Introduction

Inductive datatypes, like natural numbers, lists and trees, are very commonly used by mathematicians and programmers. One can specify such types by providing *base cases* (the numeral "0" in the case of natural numbers, the empty list – or NIL – for lists), and *inductive rules* to introduce new elements of the type, given other members of that type already exist (respectively, the functions SUC : num → num, and CONS : $\alpha \to \alpha\,\text{list} \to \alpha\,\text{list}$). Reasoning on inductive types is carried by means of mathematical induction and structural recursion.

Increasingly popular are *co-inductive datatypes*, which are instead characterized by co-induction and co-recursion. The origin of co-induction is in bisimulation, a particular case of co-induction; bisimulation originated independently in computer science, philosophical logic and set theory, and it is the most studied equality on processes in concurrency theory [10]. Co-induction is now used in many areas of computer science, like in functional languages, types, databases, and verifications tools.

Since inductive and co-inductive datatypes are so convenient, they should be easily available in programming languages, and in particular in *proof assistants*, the software systems which help scientists formalize their theorems. The possibility to define datatypes and co-datatypes would allow mathematicians and computer scientists to easily characterize (in)finite structures and reason about them.

In higher-order logic it is possible to define (co-)inductive types, but it is a tedious and repetitive task for the user, hence the necessity for a library to perform the necessary steps in a mechanical way. In this paper we will discuss a way to automatize the definition of (co-)datatypes in the HOL system.

The work is structured as follows: in 'Background' we introduce the HOL system and how users can define datatypes in it. Then, we discuss our approach based on the three functions 'all', 'map', and 'index'. In 'Example' we apply the framework to the definition of a specific co-inductive datatype. We will finally outline further work, and conclude.

Background

The HOL system HOL is a proof assistant based on higher-order logic: by higher-order logic we mean Church's *simple type theory*, augmented by the *axiom of infinity*, the *axiom of choice* (Hilbert's epsilon), and *ML-style polymorphism* (quantification on types allowed only at the outermost position). For an overview of HOL4, see for example [12].

ML is both the programming language which implements the HOL system, and the language with which the user manipulates sequents in higher-order logic (represented by objects of type `thm`).

The deductive system of HOL is based on eight rules of inference: assumption introduction, reflexivity, beta-conversion, substitution, abstraction, type instantiation, discharging an assumption, modus ponens.

Types in HOL Let the variables α, β and γ range over arbitrary types. Types can be *atomic* or *compound.* Atomic types, like unit (the type with the only element '()'), bool (the type with the two elements 'true' and 'false') and num (the type of natural numbers) are types with no arguments. Type expressions with arguments are called compound types, and they have the form $(\alpha_1, \ldots, \alpha_n)\,\mathrm{T}$ (in short $\boldsymbol{\alpha}\,\mathrm{T}$) where T is a type constructor and the α's are the arguments. Examples of type constructors are the sum type $(\alpha, \beta)\,\mathrm{sum}$ (usually denoted by $\alpha + \beta$), the product $(\alpha, \beta)\,\mathrm{prod}$ (denoted by $\alpha \times \beta$) and the function space $(\alpha, \beta)\,\mathrm{fun}$ (denoted by $\alpha \rightarrow \beta$).

Types in HOL can be obtained by the arbitrary composition of type constructors (for example, $(\alpha + \beta)\,\mathrm{list}$ or $(\alpha\,\mathrm{list})\,\mathrm{list} \rightarrow \mathrm{num}$), or by *separation*, i.e. taking a non-empty subset of an existing type. In the latter case, these are the necessary steps to carry the new type definition:

- Specify an existing type $(\alpha_1, \ldots, \alpha_n)\,\mathrm{ty}$, called the *representing type*;
- Specify a subset of this type by means of a closed term P of type $\boldsymbol{\alpha}\,\mathrm{ty} \rightarrow \mathrm{bool}$, called the *characteristic function*;
- Prove that the subset is non-empty, that is prove $\vdash \exists x.\ P\ x$;
- Introduce an axiom expressing that the new type is isomorphic to the subset of $\boldsymbol{\alpha}\,\mathrm{ty}$ specified by P.

The definitional tradition HOL follows the definitional approach: new datatypes are not generated by postulating arbitrary axioms, but rather as subsets of previously existing types. This way the logical kernel is more secure, since it does not allow extension by unsafe axioms that could introduce inconsistencies.

The two ways of creating new types above are basic and safe, but too low-level; an additional package is necessary to allow more abstract specifications of datatypes, like:

$$\text{Datatype } \alpha\,\text{list} = \textsc{nil} \mid \textsc{cons}\ \alpha\ (\alpha\,\text{list}).$$

The `datatype` package translates the specification of the inductive type above in a low-level HOL type definition, leveraging the user of many repetitive definition steps.

The original datatype package [7] used a manually defined type of finite labelled trees as a generic representing type: new datatypes were created as subsets of that type.

This was later extended by Elsa Gunter ([3] and [4]) to mutually recursive datatypes, i.e. types which depend recursively on each others, like the alternating:

$$\begin{array}{ll}\text{Datatype } (\alpha,\beta)\,\text{even} = \textsc{more}\ \alpha\ (\alpha,\beta)\,\text{odd} \mid \textsc{end}\ \alpha, \\ \quad\text{and} \quad\ (\alpha,\beta)\,\text{odd}\ = \textsc{more}\ \beta\ (\alpha,\beta)\,\text{even}\,.\end{array}$$

This approach does not efficiently handle previously defined datatypes: for example, in the inductive definition of finitely-branching trees

$$\text{Datatype } \alpha\,\text{tree} = \textsc{leaf} \mid \textsc{node}\ (\alpha\,\text{tree})\,\text{list}$$

the definition of 'list' must be first unfolded in the following way

$$\begin{array}{lll}\text{Datatype} & \alpha\,\text{treelist} & = \textsc{nil} \mid \textsc{cons}\ (\alpha\,\text{tree})\ (\alpha\,\text{treelist}), \\ \text{and} & \alpha\,\text{tree} & = \textsc{leaf} \mid \textsc{node}\ \alpha\,\text{treelist}\end{array}$$

leading to repeated definitions and bad performances in case of many nestings.

Another issue is the lack of support for co-inductive datatypes, like *lazy lists*:

$$\text{Co-datatype } \alpha\,\text{llist} = \textsc{lcons}\ \alpha\,\text{llist}\,.$$

Lazy lists or *streams* model infinite sequences, and their specification differs from the one of usual lists in the fact that the first have no base case (no empty lazy list). If characterized inductively, the specification of lazy lists would yield the empty type, while co-inductively the type contains only "infinite lists" like for example:

$$\text{Co-inductive zeroes} \overset{\text{def}}{=} \textsc{lcons}\ 0\ \text{zeroes}\,.$$

Some works in the literature overcome the problems we just outlined (no co-inductive datatypes, no efficient re-using of previously defined datatypes): our main inspiration is the approach in [13] for datatypes in Isabelle/HOL. It is based on category theory, and it achieves modularity and compositionality by making use of *rich type constructors*, which are type constructors satisfying certain properties justified by categorical operations.

Our approach draws heavily on that work, in that it equips types with additional functions, which allow one to reuse previously defined datatypes instead of replaying their definition. Nevertheless, our approach mostly follows the tradition by means of concrete constructions with trees. Its additional strenghts are: it does not depend on a theory of cardinals, nor on the axiom of choice, thus it is completely constructive; our functors are in a nice continuation-passing style, have a uniform format, and they are already defined in the HOL library.

Our framework: map, all, index

Let $\boldsymbol{\alpha}\,\mathrm{T}$ be an n-ary type constructor. We equip every such type constructor with the following three functions:

$$\begin{aligned}\mathrm{map} &: [\alpha_i \to \beta_i]_i \qquad\; \to (\boldsymbol{\alpha}\,\mathrm{T} \to \boldsymbol{\beta}\,\mathrm{T})\\ \mathrm{index} &: [\alpha_i \to \iota_i \to \gamma]_i \to (\boldsymbol{\alpha}\,\mathrm{T} \to \boldsymbol{\iota}\,\mathrm{T}_{\mathrm{index}} \to \gamma)\\ \mathrm{all} &: [\alpha_i \to \mathrm{bool}]_i \quad\; \to (\boldsymbol{\alpha}\,\mathrm{T} \to \mathrm{bool})\end{aligned}$$

The square brackets in the type signatures above are just syntactic abbreviations. For example, the signature for the map function for a generic n-ary type constructor T is:

$$(\alpha_1 \to \beta_1) \to \ldots \to (\alpha_n \to \beta_n) \to (\boldsymbol{\alpha}\,\mathrm{T} \to \boldsymbol{\beta}\,\mathrm{T}).$$

The type signatures of the functions above share the same structure: they all start requiring n continuations of a certain format (one for each type variable) and return a “lifted” function, with signature of a similar format but now talking of the type constructor T.

The usual type of lists with elements of type α (denoted by $\alpha\,\mathrm{list}$) will serve to illustrate the following concepts. It will be of help to visualize the elements of a datatype $\boldsymbol{\alpha}\,\mathrm{T}$ as *shape* plus *content*. Let’s consider a list, say $\Big[\boxed{1},\boxed{2},\boxed{3}\Big]$: the structure of this list is $[\,\cdot\,,\,\cdot\,,\,\cdot\,]$, while the atoms $\boxed{1}$, $\boxed{2}$ and $\boxed{3}$ are the content.

We will now describe the intuition behind each function. For examples of these functions for known type constructors, see the appendix.

Map function The map function (known in category theory and present in functional programming languages) maps the type variables underlying a type constructor. Mainly known for its version on lists:

$$\mathrm{map}_{\mathrm{list}} : (\alpha \to \beta) \to \alpha\,\mathrm{list} \to \beta\,\mathrm{list},$$

The result of $\mathrm{map}\, f\, [x, y, z]$ is the list obtained by applying f to each element of the list, that is $[f\ x, f\ y, f\ z]$. By mapping a function to a list, we do not modify the shape of the given list; instead we leave the structure untouched and we modify the elements of the list in their respective positions. Generalizing this intuition, we can say that map functions do not rearrange the structure of input items, instead they apply functions on their atoms “in place”.

Index function We already know how the 'index' function works for lists: it takes a list and a natural number and returns the element of that list at the specified position. Its signature is:

$$\text{index } : \alpha \,\text{list} \to \mathbb{N} \to \alpha.$$

This function is not general enough for our purposes: α could be a composite type itself, and what stops us from wanting to keep digging into it, in order to retrieve deeper atoms? Let's now change slightly the signature above: we can write

$$\text{index}_{\text{list}} : (\ldots) \to \alpha \,\text{list} \to \mathbb{N} \times \ldots \to \alpha.$$

The index function in the continuation-passing style will have type:

$$\text{index}_{\text{list}} : (\alpha \to \iota \to \gamma) \to \alpha \,\text{list} \to \mathbb{N} \times \iota \to \gamma.$$

The difference between index and $\text{index}_{\text{list}}$ is that the latter additionally takes in input a function of type $\alpha \to \iota \to \gamma$, in order to continue the retrieval on the element of the list. But to continue retrieving, we need an additional "cursor" ι as well. The implementation is:

$$\text{index}_{\text{list}}\ g\, l\, (n, i) = g\, (\text{index}\, l\, n)\, i.$$

Summing up, index functions take in input a "coordinate" inside an item, and operate on the atom at that position, ignoring the other atoms and the rest of the structure.

All function This function is also known as "EVERY" in the HOL system. Given a predicate, the function EVERY for lists checks whether this predicate holds on every element of a given list:

$$\begin{aligned}
\text{all}_{\text{list}}\, P\ \textsc{nil} &= \text{true} \\
\text{all}_{\text{list}}\, P\ (\textsc{cons}\ a\ as) &= P\ a \wedge \text{all}\, P\ as
\end{aligned}$$

In general, "all" functions take n predicates and given an item they check whether all its atoms satisfy the corresponding predicate.

Required properties In order to ensure correctness of the datatype definition, the three functions above should satisfy certain properties:

- respect composition (parametricity for map)

$$\begin{aligned}
(\text{all}\, P_1 \ldots P_n) \circ (\text{map}\, f_1 \ldots f_n) &= \quad \text{all}\ (P_1 \circ f_1) \ldots (P_n \circ f_n) \\
(\text{index}\, g_1 \ldots g_n) \circ (\text{map}\, f_1 \ldots f_n) &= \text{index}\ \ (g_1 \circ f_1) \ldots (g_n \circ f_n) \\
(\text{map}\, f'_1 \ldots f'_n) \circ (\text{map}\, f_1 \ldots f_n) &= \ \text{map}\ \ (f'_1 \circ f_1) \ldots (f'_n \circ f_n)
\end{aligned}$$

- identity for all

$$\text{all}\, (\text{K}\, \text{true}) \cdots (\text{K}\, \text{true}) = \text{K}\, \text{true}$$

- map and index should be injective together. For every h and k define $\mathrm{Inj}(h,k)$ iff the function $\lambda x.\,\langle h(x), k(x)\rangle$ is injective.
  We require the following

$$\mathrm{Inj}(f_1, g_1)\ldots, \mathrm{Inj}(f_n, g_n) \Rightarrow \mathrm{Inj}(\mathrm{map}\ f_1 \ldots f_n\ ,\ \mathrm{index}\ g_1 \ldots g_n).$$

The above requirements (which are just part of the complete list of properties) ara necessary later for the construction of new types: the last condition to prove injectivity of the constructor, the first two to prove the datatype nonempty.

Example: forking process

As an example for the use of our framework, let us define the following codatatype:

$$\alpha\,\mathrm{process} \overset{\mathsf{codef}}{=} \textsc{return}\ \alpha \mid \textsc{fork}\ (\mathrm{process\,list}).$$

The motivation for this example is a process in a computer which can either terminate yielding a value of type α, or fork itself and execute a finite number of other processes. The co-inductive definition allows to have processes which could not terminate, intuitively obtained by an infinite application of the FORK constructor.

From a semantical point of view, in order to carry out the datatype definition we need to solve the fixpoint equation $\mathrm{ty} \approx \alpha + \mathrm{ty\,list}$ in a "maximal way", that is we take the *greatest fixpoint* of the operator $(\alpha, \mathrm{ty})F \overset{\mathsf{def}}{=} \alpha + (\mathrm{ty\,list})$ w.r.t. the variable ty.

Our goal is to translate the type specification above to a type definition in HOL. The main steps will be:

1. use as representing type (an instance of) the type of polymorphic labelled trees with infinite depth and possibly infinite branching;
2. define by co-induction a subset of the representing type;
3. provide a witness for the new datatype.

Representing type In order to find the correct representing type, we use the three functions which we introduced in the sections above.

First we construct the functions 'all', 'map' and 'index' for the type operator F by composing the corresponding functions for 'sum' and 'list':

$$\begin{aligned}
\mathrm{all_F}\ P_\alpha\ P_{\mathrm{ty}} &\overset{\mathsf{def}}{=} \mathrm{all_{sum}}\quad P_\alpha\ (\mathrm{all_{list}}\quad P_{\mathrm{ty}}\ ),\\
\mathrm{map_F}\ f_\alpha\ f_{\mathrm{ty}} &\overset{\mathsf{def}}{=} \mathrm{map_{sum}}\ \ f_\alpha\ (\mathrm{map_{list}}\ \ f_{\mathrm{ty}}\ ),\\
\mathrm{index_F}\ g_\alpha\ g_{\mathrm{ty}} &\overset{\mathsf{def}}{=} \mathrm{index_{sum}}\ g_\alpha\ (\mathrm{index_{list}}\ g_{\mathrm{ty}}\ ).
\end{aligned}$$

The type signatures are:

$$\begin{aligned}
\mathrm{all_F} &: (\alpha \to \mathrm{bool}) \to (\mathrm{ty} \to \mathrm{bool}) \to (\alpha, \mathrm{ty})\,\mathrm{F} \to \mathrm{bool},\\
\mathrm{map_F} &: (\alpha \to \alpha') \to (\mathrm{ty} \to \mathrm{ty}') \to (\alpha, \mathrm{ty})\,\mathrm{F} \to (\alpha', \mathrm{ty}')\,\mathrm{F},\\
\mathrm{index_F} &: (\alpha \to \iota_\alpha \to \gamma) \to (\mathrm{ty} \to \iota_{\mathrm{ty}} \to \gamma) \to (\alpha, \mathrm{ty})\,\mathrm{F} \to (\iota_\alpha \times \mathrm{num} \times \iota_{\mathrm{ty}}) \to \gamma.
\end{aligned}$$

Since we aim at representing the 'process' type in a type of trees, we store the "non-recursive" part of the elements in labels of the tree (which have type $\alpha + \mathrm{unit\,list}$), while we store the "recursive" part in the children. For example, in the element $\textsc{return}\ a$ do not occur atoms of type 'process', thus we represent it with the tree with a single node labelled $\textsc{inl}\ a$ and no children. Suppose now we have $\textsc{fork}\ [\textsc{return}\ a_1, \textsc{return}\ a_2, \textsc{fork}\ [\textsc{return}\ a_3]]$: we store its shape $\textsc{inr}\ [(), (), ()]$ in the label of the root node, and then we store recursively the three elements $\textsc{return}\ a_1$, $\textsc{return}\ a_2$, and $\textsc{fork}\ [\textsc{return}\ a_3]$ each in one child.

Let's generalize the argument above: for $x\colon (\alpha, \mathrm{ty})F$, the shape of x is given by erasing all its atoms of type ty (mapping them to $()\colon$ unit), while its content are the atoms of type ty that can be indexed:

$$\begin{aligned} \text{"shape of } x\text{"} &\stackrel{\mathsf{def}}{=} (\mathrm{map}_{\mathrm{F}} \quad \mathrm{I} \quad \textsc{arb})\ x \quad : (\alpha, \mathrm{unit})\,\mathrm{F} \\ \text{"atom of } x \text{ at } p\text{"} &\stackrel{\mathsf{def}}{=} (\mathrm{index}_{\mathrm{F}}\ \textsc{arb} \quad \mathrm{K}\ \ )\ x\ p\ : \mathrm{ty}, \end{aligned}$$

where:

- the term 'I' on the first line is the identity function on α (we leave the atoms of type α in x untouched);
- the term ARB in the first line is an arbitrary function of type ty → unit, which can be only the constant function mapping everything to $()$;
- the term ARB on the second line is an arbitrary function of type $\alpha \to \mathrm{unit} \to \mathrm{ty}$ (we are not interested in indexing in x atoms of type α, but only those of type ty);
- the term $\mathrm{K}\colon \mathrm{ty} \to \mathrm{unit} \to \mathrm{ty}$ on the second line is $\lambda x \lambda y.\ x$.

We are now going to apply the fixpoint operation: we first define $C\colon (\alpha, \mathrm{ty})\,\mathrm{F} \to \mathrm{ty}$, the constructor for the datatype specified by F. Our goal is to provide ty as well, and it will be a consequence of the definition of C. Recall that we want to host our datatype on a tree type: this is why we force ty to have the form "(branch) list → label". We can thus define $C\ x$ by cases on lists:

$$\begin{aligned} C\ x\ \textsc{nil} \quad\quad &\stackrel{\mathsf{def}}{=} \text{"shape of } x\text{"} \\ C\ x\ (\textsc{cons}\ p\ ps) &\stackrel{\mathsf{def}}{=} (\text{"atom of } x \text{ at } p\text{"})\ ps \end{aligned}$$

By the definition above, the type of C uniquely determines the types branch$\stackrel{\mathsf{def}}{=}$, and ty:

$$\mathrm{ty} \stackrel{\mathsf{def}}{=} (\mathrm{unit} \times \mathrm{num} \times \mathrm{unit})\,\mathrm{list} \to (\alpha, \mathrm{unit})\,\mathrm{F}\,.$$

We now have the constructor, and we can prove useful properties like injectivity and distinctness (for the split version of the constructor).

Characteristic function Once found the hosting type ty and the constructor C, the next step is to carve out of ty a co-inductively characterized subset. In our case, we define the predicate P by the following co-inductive rule:

$$\frac{\mathrm{all}_{\mathrm{F}}\ (\mathrm{K}\,\mathrm{true})\ P\ x}{P\ (C\ x)}$$

The meaning for ‘K true’ is that at this point we don’t distinguish atoms of type α: we are only interested in those of type ty, for which we check if P holds. The intuition behind the rule above is: whenever P holds for all the atoms of type ty occurring in x, then P holds for $C\ x$ as well. Since we characterize P co-inductively, we take the biggest P satisfying the rule above: this is possible in HOL through the `CoIndDef` library, which allows to define predicates characterized by co-induction. This library was implemented by the author by adapting the original `IndDef` package for inductive definitions originally by Harrison and Melham (see [5]).

Prove non-emptiness Before concluding the definition of the new datatype, we have to prove that it is non-empty; in our case, this means finding a term t such that $C\ t$ holds. Since we have a base case, this step is simple: take for example $t \overset{\text{def}}{=} \textsc{inl arb}^\alpha$ (a return value); it is easy to see that P holds on $C\ t$ from the rule above, since all_F (K true) P ($\textsc{inl arb}$) $\equiv$ (K true) $\textsc{arb}$ $\equiv$ true. C ($\textsc{inl arb}$) stands for ‘$\textsc{return arb}$’.

Type definition Now that we have a witness for the non-emptiness of P, we have a proof for $\vdash \exists x.\ P\ x$. We can call the HOL primitive to define a new datatype:

$$\texttt{new\_type\_definition}\ (\vdash \exists x.\, P\, x)$$

The process datatype is defined, but there would be many other actions to carry: the current HOL datatype package proves many useful theorems about the newly defined datatype. For example, the initiality/finality theorem, structural (co-)induction theorem, the case analysis theorem, etc.

Conclusion and Further Work

We presented a framework to define types in HOL. The framework handles both datatypes and co-datatypes, and supports type specifications in which occur types equipped with three functions: ‘map’, ‘index’ and ‘all’. These functions have a clear meaning, and are well-known for common types. Unlike the existing package, our framework allows defining co-datatypes too, and efficiently handles nested and mutual definitions.

The draft for a new `datatype` package based on the framework in this paper can be found at:

https://github.com/HOL-Theorem-Prover/HOL/tree/master/src/new-datatype,

but it is still at an early stage; we implemented the composition of the properties for new types, the construction of non-emptiness witnesses, and the injectivity of the constructor, but many other results are missing. In fact, the current `datatype` package proves for every new type the following useful theorems: initiality theorem, injectivity of the constructors, distinctness of the constructors, structural induction, case analysis, definition of the 'case' constant for the type, congruence for the case constant, definition of the 'size' of the type. Since these theorems are crucial for the user, it is essential for the new package to implement them too.

Acknowledgment: This started as a research project at National ICT Australia in Canberra, under the supervision of Michael Norrish, which I thank for guidance and fruitful discussions. I was supported by the European Master's Program in Computational Logic (EMCL).

References

1. Proceedings of the 27th Annual IEEE Symposium on Logic in Computer Science, LICS 2012, Dubrovnik, Croatia, June 25-28, 2012. IEEE Computer Society (2012)
2. Claesen, L.J.M., Gordon, M.J.C. (eds.): Higher Order Logic Theorem Proving and its Applications, Proceedings of the IFIP TC10/WG10.2 Workshop HOL'92, Leuven, Belgium, 21-24 September 1992, IFIP Transactions, vol. A-20. North-Holland/Elsevier (1993)
3. Gunter, E.L.: Why we can't have SML-style datatype declarations in HOL. In: Claesen and Gordon [2], pp. 561–568
4. Gunter, E.L.: A broader class of trees for recursive type definitions for HOL. In: Joyce and Seger [6], pp. 141–154
5. Harrison, J.: Inductive definitions: Automation and application. In: Schubert et al. [11], pp. 200–213
6. Joyce, J.J., Seger, C.H. (eds.): Higher Order Logic Theorem Proving and its Applications, 6th International Workshop, HUG '93, Vancouver, BC, Canada, August 11-13, 1993, Proceedings, Lecture Notes in Computer Science, vol. 780. Springer (1994)
7. Melham, T.F.: Current trends in hardware verification and automated theorem proving. chap. Automating Recursive Type Definitions in Higher Order Logic, pp. 341–386. Springer-Verlag New York, Inc., New York, NY, USA (1989)
8. Mohamed, O.A., Muñoz, C.A., Tahar, S. (eds.): Theorem Proving in Higher Order Logics, 21st International Conference, TPHOLs 2008, Montreal, Canada, August 18-21, 2008. Proceedings, Lecture Notes in Computer Science, vol. 5170. Springer (2008)
9. Paulson, L.C.: Mechanizing coinduction and corecursion in higher-order logic. J. Log. Comput. 7(2), 175–204 (1997)
10. Sangiorgi, D.: On the origins of bisimulation and coinduction. ACM Trans. Program. Lang. Syst. 31(4) (2009)
11. Schubert, E.T., Windley, P.J., Alves-Foss, J. (eds.): Higher Order Logic Theorem Proving and Its Applications, 8th International Workshop, Aspen Grove, UT, USA, September 11-14, 1995, Proceedings, Lecture Notes in Computer Science, vol. 971. Springer (1995)
12. Slind, K., Norrish, M.: A brief overview of HOL4. In: Mohamed et al. [8], pp. 28–32
13. Traytel, D., Popescu, A., Blanchette, J.C.: Foundational, compositional (co)datatypes for higher-order logic: Category theory applied to theorem proving. In: Proceedings of the 27th Annual IEEE Symposium on Logic in Computer Science, LICS 2012, Dubrovnik, Croatia, June 25-28, 2012 [1], pp. 596–605

Appendix

sum type

$\text{all}_{\text{sum}} : \left[\begin{smallmatrix}\alpha_1 \to \text{bool}\\ \alpha_2 \to \text{bool}\end{smallmatrix}\right] \to (\alpha_1 + \alpha_2) \to \text{bool}$
$\text{all}_{\text{sum}}\ P_1\ P_2\ (\textsc{inl}\ a_1) = P_1\ a_1$
$\text{all}_{\text{sum}}\ P_1\ P_2\ (\textsc{inr}\ a_2) = P_2\ a_2$

$\text{map}_{\text{sum}} : \left[\begin{smallmatrix}\alpha_1 \to \beta_1\\ \alpha_2 \to \beta_2\end{smallmatrix}\right] \to (\alpha_1 + \alpha_2) \to (\beta_1 + \beta_2)$
$\text{map}_{\text{sum}}\ f_1\ f_2\ (\textsc{inl}\ a_1) = \textsc{inl}\ (f_1\ a_1)$
$\text{map}_{\text{sum}}\ f_1\ f_2\ (\textsc{inr}\ a_2) = \textsc{inr}\ (f_2\ a_2)$

$\text{index}_{\text{sum}} : \left[\begin{smallmatrix}\alpha_1 \to \iota_1 \to \gamma\\ \alpha_2 \to \iota_2 \to \gamma\end{smallmatrix}\right] \to (\alpha_1 + \alpha_2) \to (\iota_1 \times \iota_2) \to \gamma$
$\text{index}_{\text{sum}}\ g_1\ g_2\ (\textsc{inl}\ a_1)\ \langle i_1, i_2\rangle = g_1\ a_1\ i_1$
$\text{index}_{\text{sum}}\ g_1\ g_2\ (\textsc{inr}\ a_2)\ \langle i_1, i_2\rangle = g_2\ a_2\ i_2$

prod type

$\text{all}_{\text{prod}} : \left[\begin{smallmatrix}\alpha_1 \to \text{bool}\\ \alpha_2 \to \text{bool}\end{smallmatrix}\right] \to (\alpha_1 \times \alpha_2) \to \text{bool}$
$\text{all}_{\text{prod}}\ P_1\ P_2\ \langle a_1, a_2\rangle = P_1\ a_1 \wedge P_2\ a_2$

$\text{map}_{\text{prod}} : \left[\begin{smallmatrix}\alpha_1 \to \beta_1\\ \alpha_2 \to \beta_2\end{smallmatrix}\right] \to (\alpha_1 \times \alpha_2) \to (\beta_1 \times \beta_2)$
$\text{map}_{\text{prod}}\ f_1\ f_2\ \langle a_1, a_2\rangle = \langle f_1\ a_1, f_2\ a_2\rangle$

$\text{index}_{\text{prod}} : \left[\begin{smallmatrix}\alpha_1 \to \iota_1 \to \gamma\\ \alpha_2 \to \iota_2 \to \gamma\end{smallmatrix}\right] \to (\alpha_1 \times \alpha_2) \to (\iota_1 + \iota_2) \to \gamma$
$\text{index}_{\text{prod}}\ g_1\ g_2\ \langle a_1, a_2\rangle\ (\textsc{inl}\ i_1) = g_1\ a_1\ i_1$
$\text{index}_{\text{prod}}\ g_1\ g_2\ \langle a_1, a_2\rangle\ (\textsc{inr}\ i_2) = g_2\ a_2\ i_2$

fun type

$\text{all}_{\text{fun}} : (\beta \to \text{bool}) \to (\alpha \to \beta) \to \text{bool}$
$\text{all}_{\text{fun}}\ P\ f = \forall x.\ P(f\ x)$

$\text{map}_{\text{fun}} : (\beta \to \delta) \to (\alpha \to \beta) \to (\alpha \to \delta)$
$\text{map}_{\text{fun}} = \circ$ (function composition)

$\text{index}_{\text{fun}} : (\beta \to \iota \to \gamma) \to (\alpha \to \beta) \to \alpha \times \iota \to \gamma$
$\text{index}_{\text{fun}}\ g\ f\ (a, i) = g\ (f\ a)\ i$

Fig. 1. all, map, index functions for basic type constructors

option type

$\mathrm{all}_{\mathrm{option}} : (\alpha \to \mathrm{bool}) \to (\alpha\,\mathrm{option}) \to \mathrm{bool}$
$\mathrm{all}_{\mathrm{option}}\, P\, (\textsc{some}\ a) = P\ a$
$\mathrm{all}_{\mathrm{option}}\, P\ \ \textsc{none}\ \ \ = \mathrm{true}$

$\mathrm{map}_{\mathrm{option}} : (\alpha \to \beta) \to \alpha\,\mathrm{option} \to \beta\,\mathrm{option}$
$\mathrm{map}_{\mathrm{option}}\, f\ (\textsc{some}\ a) = \textsc{some}\ (f\ a)$
$\mathrm{map}_{\mathrm{option}}\, f\ \ \textsc{none} = \textsc{none}$

$\mathrm{index}_{\mathrm{option}} : (\alpha \to \iota \to \gamma) \to \alpha\,\mathrm{option} \to \iota \to \gamma$
$\mathrm{index}_{\mathrm{option}}\, g\ (\textsc{some}\ a)\ i = g\ a\ i$

list type

$\mathrm{all}_{\mathrm{list}} : (\alpha \to \mathrm{bool}) \to \alpha\,\mathrm{list} \to \mathrm{bool}$
$\mathrm{all}_{\mathrm{list}}\, P\ \textsc{nil}\ = \mathrm{true}$
$\mathrm{all}_{\mathrm{list}}\, P\ (\textsc{cons}\ a\ as) = P\ a \wedge \mathrm{all}_{\mathrm{list}}\, P\ as$

$\mathrm{map}_{\mathrm{list}} : (\alpha \to \beta) \to \alpha\,\mathrm{list} \to \beta\,\mathrm{list}$
$\mathrm{map}_{\mathrm{list}}\, f\ \textsc{nil} = \textsc{nil}$
$\mathrm{map}_{\mathrm{list}}\, f\ (\textsc{cons}\ a\ as) = \textsc{cons}\ \ (f\ a)\ (\mathrm{map}_{\mathrm{list}}\, f\ as)$

$\mathrm{index}_{\mathrm{list}} : (\alpha \to \iota \to \gamma) \to \alpha\,\mathrm{list} \to (\mathrm{num} \times \iota) \to \gamma$
$\mathrm{index}_{\mathrm{list}}\, g\ (\textsc{cons}\ a\ as)\ \langle 0, i\rangle = g\ a\ i$
$\mathrm{index}_{\mathrm{list}}\, g\ (\textsc{cons}\ a\ as)\ \langle \mathrm{S}\, n, i\rangle = \mathrm{index}_{\mathrm{list}}\, g\ as\ (n, i)$

atomic type

$\mathrm{all}_{\mathrm{num}} : \mathrm{num} \to \mathrm{bool}$
$\mathrm{all}_{\mathrm{num}} = \mathrm{K}\,\mathrm{true}$ (constantly true)

$\mathrm{map}_{\mathrm{num}} : \mathrm{num} \to \mathrm{num}$
$\mathrm{map}_{\mathrm{num}} = \mathrm{I}$ (identity)

$\mathrm{index}_{\mathrm{num}} : \mathrm{num} \to \mathrm{unit} \to \gamma$
$\mathrm{index}_{\mathrm{num}}$ undefined (arbitrary)

Fig. 2. all, map, index functions for common type constructors

Grades of Responsibility

Estelle Doriot

Utrecht University

Abstract. In this paper, we give a formal definition of grades of responsibility using stit logic and probabilities. We formalize the notion of responsibility based on probability increase. Then, we compare our framework with similar notions developed by Braham and van Hees.

1 Introduction

The question of responsibility for the outcome of one's choices is now relevant to the study of intelligent systems as those are increasingly integrated to our daily life. The framework we present here reflects the notion of agent-responsibility as defined by [8] which says that an individual is agent-responsible for an outcome to the extent that it suitably reflects the exercise of his agency. This notion is more general than moral responsibility as it only requires that the agent has the ability to make an autonomous choice. For example, we may consider that a child is not morally responsible for a crime he committed because he is not morally aware; but he would be agent responsible. Thus, there is no deontic component in our definition of responsibility, but the addition of a deontic operator could be a future research topic. To be more precise, we are interested in representing grades of responsibility as we believe that responsibility is not an all or nothing notion [8]. How to represent responsibility in relation to agents, causal relations and probabilities is still an open question. We formalize the notion of responsibility based on probability increase.

To that end, we extend stit theory, which is a formal theory of agency well established in philosophy [1]. More recently, stit theory has been used for the specification of multi-agent systems [6]. Stit models the agents' choices; however, the final outcome is not known by the agents before the execution of the action because it is the result of the combined choices and actions of all the participants. Thus an agent's action might not be successful due to other agents – or the environment – counteracting him. The responsibility of an agent depends on his beliefs about the outcome of his choice. That is why we use probabilities to represent the subjective beliefs of an agent about the combined choices of the other agents. The responsibility of one agent's choice for a particular outcome is defined as the increased probability of this outcome happening due to the agent's choice, which measures the contribution of the agent to the outcome. This definition of responsibility is subjective due to the use of subjective probabilities. We suppose that each agent consciously makes a choice and knows what choice he makes.

After giving the syntax and the semantics of an object language that represents grades of responsibility, we compare it with the notions of responsibility and degree of causation defined by Braham and van Hees in [3] and [4]. Even though theses notions are not similar to our definition of grades of responsibility, they are the closest we have found in the literature.

2 Syntax

We introduce here the syntax of our logic, which is an extension of the syntax of the probabilistic XSTIT logic defined in [5]. A new responsibility operator has been added to the existing operators. Moreover, there are differences in the semantics: we use a *Choices* function instead of an h-effectivity function, and the definition of subjective probability varies as well.

Definition 1. *Given a countable set of propositions P, a finite set $\mathcal{A}$ of agent names and a set of real numbers $C \subseteq [0,1]$, the formal language $\mathcal{L}$ is:*

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid X\varphi \mid [\alpha\ \texttt{xstit}^{\geq c}]\varphi \mid [\alpha\ \texttt{resp}^{=c}]\varphi$$

where $p \in P$, $\alpha \in \mathcal{A}$ and $c \in C$.

The modal operators have the following interpretation: $\Box\varphi$ expresses 'historical necessity'. We abbreviate $\neg\Box\neg\varphi$ by $\Diamond\varphi$; $X\varphi$ has a standard interpretation as the transition to a next moment; $[\alpha\ \texttt{xstit}^{\geq c}]\varphi$ stands for 'agent α sees to it that φ in the next moment with a probability of at least c' and $[\alpha\ \texttt{resp}^{=c}]\varphi$ stands for 'agent α has a responsibility of c for φ in the next moment'.

3 Semantics

The underlying structure of the frame we use is a branching-time structure made up of moments and histories going through these moments.

Definition 2 (History). *Given M a non-empty set of* moments*, H is a non-empty set of possible system* histories *isomorphic with $\ldots, m_{-2}, m_{-1}, m_0, m_1, m_2, \ldots$ with $m_x \in M$ for $x \in \mathbb{Z}$. $H_m = \{h \in H \mid m \in h\}$ is the set of histories going through m. We define two functions succ and prec by: $m' = succ(m,h)$ iff m' succeeds m on the history h, and $m = prec(m',h)$ iff m precedes m' on the history h. We have the following constraints on the set H:*
(One successor) *$\forall m, h$, if $m \in h$ then $\exists! m' = succ(m,h)$*
(Definition predecessor) *$m = prec(m',h)$ iff $m' = succ(m,h)$*
(One predecessor) *$\forall m, h$, if $m \in h$ then $\exists! m' = prec(m,h)$*
(One past) *$\forall m, \forall h, h' \in H_m$, $prec(m,h) = prec(m,h')$*

Condition **(One successor)** states that there is one and only one successor of a moment m along a history h going through m; condition **(definition predecessor)** is the definition of *prec* as the converse of *succ*; condition **(One predecessor)** states that there is one and only one predecessor of a moment m along a history h going through m; and condition **(One past)** expresses the fact that there is one past.

Definition 3 (Choices). *Choices*: $M \times \mathcal{A} \to \wp(\wp(H))$ *is the* choices function *yielding for an agent α in a moment m the set $Choices(m,\alpha)$ of subsets of H_m containing the agent's choices. We have the following constraints on $Choices(m,\alpha)$:*

(No empty choice) $\forall K \in Choices(m,\alpha),\ K \neq \emptyset$
(No absence of choice) $\forall h \in H_m,\ \exists K \in Choices(m,\alpha)$ *such that* $h \in K$
(No choice between undivided histories) $\forall h, h'$, *if*
$succ(m,h) = succ(m,h')$ *then* $\forall K \in Choices(m,\alpha)$, *if* $h \in K$ *then* $h' \in K$
(Independence of agency) $\forall \alpha_i \in \mathcal{A}, \forall K_i \in Choices(m,\alpha_i), \bigcap_{\alpha_i \in \mathcal{A}} K_i \neq \emptyset$

The choice made by agent α in a moment m relative to the history h is given by: $Choice(m,\alpha,h) = \{\bigcup_i K_i \mid K_i \in Choices(m,\alpha)$ *and* $h \in K_i\}$.

We thus define a choice K as a set of histories; and $Choices(m,\alpha)$ is the set of all the choices available to agent α at m. It is equivalent to an effectivity function. Condition **(No empty choice)** states that a choice cannot be empty; condition **(No absence of choice)** states that every history is accessible to any agent through a choice. The following conditions are the well-known stit condition 'no choice between undivided histories' and the 'independence of agency' which states that the choice of one agent cannot limit the choices the other agents make simultaneously. The set $Choices(m,\alpha)$ is not a partition of H_m because we omit the fact that the choices in $Choices(m,\alpha)$ are mutually disjoint. We do not include this condition because it is not modally expressible and will thus have no impact on our logic.

The definition of *Choices* can be extended to groups of agents as the intersection of the choices of all the agents in this group.

Definition 4 (Group choices). $Choices_G$: $M \times \wp(\mathcal{A}) \to \wp(\wp(H))$ *is the* group choice function *yielding for a group of agents A in a moment m the set of subsets of H_m containing the group's combined choices:* $K \in Choices_G(m,A)$ *iff* $\exists(K_{\alpha_1},\ldots,K_{\alpha_k}) \in \times_{\alpha_i \in A} Choices(m,\alpha_i)$ *where* $K = \cap K_{\alpha_i}$.
The choice made by the group of agents A in a moment m along the history h is given by: $Choice_G(m,A,h) = \{\bigcup_i K_i \mid K_i \in Choices_G(m,A)$ *and* $h \in K_i\}$.

We introduce the state of an agent α at a moment m, which is the combined choices of all the other agents at that same moment m.

Definition 5 (States). *The* states *of an agent α at a moment m is defined by:* $States(m,\alpha) = Choices_G(m, \mathcal{A} \setminus \{\alpha\})$.

For each agent, the expectation is a subjective probability distribution over the states of this agent at a moment. We choose to define the expectation as a probability distribution over states (which are the combined choices of all the other agents) rather than on the individual choices of the other agents. This allows for cases where the choices of the agents are correlated.

Definition 6 (Expectation). *The* expectation *function* B: $M \times \mathcal{A} \times \wp(H) \to C$ *is a subjective probability function such that $B(m,\alpha,K)$ expresses agent α's expectation that he will be in a state K in a moment m. We apply the following constraints:*

1. $B(m, \alpha, K) \geq 0$ *if* $K \in \mathit{States}(m, \alpha)$
2. $B(m, \alpha, K) = 0$ *otherwise*
3. $\sum_{K \in \mathit{States}(m,\alpha)} B(m, \alpha, K) = 1$

Conditions **1.** and **2.** express that only states can be assigned a non-zero expectation. Condition **3.** is a standard probability condition stating that the sum of the expectation of one agent over the possible states adds up to 1. As stated in the definition of the syntax, the set C is a subset of $[0, 1]$. One might argue that the probability of a state should not be zero (and that the inequality in condition **1.** should be strict) because it would represent an impossible situation. However, we are using subjective probabilities which reflects the beliefs of the agents. This means that a zero probability would represent a situation that the agent did not know was possible.

We now have all the elements necessary to define the probabilistic XSTIT-frame: a set of moments, a set of histories, a choice function and an expectation function.

Definition 7 (Probabilistic XSTIT-frame). *A* probabilistic XSTIT-frame *is a tuple* $\mathcal{F} = \langle M, H, \mathit{Choices}, B \rangle$ *such that:*
1. M *is a non-empty set of moments*
2. H *is a non-empty set of histories*
3. $\mathit{Choices}\colon M \times \mathcal{A} \to \wp(\wp(H))$ *is a choice function*
4. $B\colon M \times \mathcal{A} \times \wp(H) \to C$ *is an expectation function*

Definition 8 (Probabilistic XSTIT-model). *A frame is extended to a* model $\mathcal{F} = \langle M, H, \mathit{Choices}, B, V \rangle$ *by adding a valuation* V *of atomic propositions* $V\colon P \to \wp(M)$ *assigning to each atomic proposition the set of moments relative to which they are true.*

Figure 1 visualizes a probabilistic XSTIT-model with two agents, as given by definition 8. The cells represents the moments and the dashed lines going through them are the histories, grouped in bundles. In each game form, the columns correspond to the choices of the first agent α_1 and the rows to the choices of the second agent α_2. The numbers present alongside the rows are the probabilities assigned by agent α_1 to the actions of agent α_2 and, symmetrically, the numbers alongside the columns are the probabilities assigned by agent α_2 to the actions of agent α_1.

We have defined the frames and models but we still need a few additional definitions to reach the notion of responsibility. First of all, we are looking at the result of the agent's actions: when does an action leads to φ? The 'possible next φ-states' function gives the set of states that, given the agent's choice, will ensure that φ is true at the next moment.

Definition 9 (Possible next φ-states). *The* possible next φ-states *function* $\mathit{PosX}\colon M \times H \times \mathcal{A} \times \mathcal{L} \to \wp(\wp(H))$ *which for a moment* m*, a history* h*, an agent* α *and a formula* φ *gives the possible next states obeying* φ *given the agent's current choice determined by* h*, is defined by:* $\mathit{PosX}(m, h, \alpha, \varphi) = \{K \in \mathit{States}(m, \alpha) \mid \forall h' \in K \cap \mathit{Choice}(m, \alpha, h), \langle \mathit{succ}(m, h'), h' \rangle \models \varphi\}$.

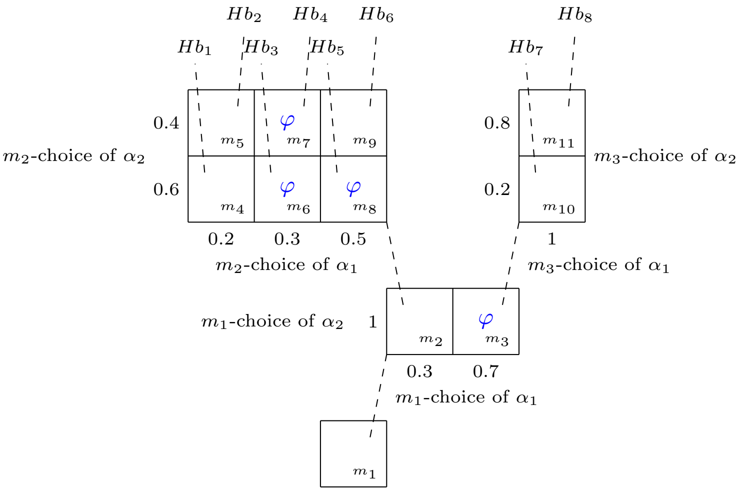


Fig. 1: A partial probabilistic XSTIT model with two agents

The chance of success of an agent's action resulting to φ is the subjective belief of this agent that the choice he makes will result in a situation where φ is true. It is defined as the sum of the probabilities assigned to the states in which, given the choice made by the agent, φ will be true in the next moment.

Definition 10 (Chance of success). *The* chance of success *function* $CoS\colon M \times H \times \mathcal{A} \times \mathcal{L} \to C$ *which for a moment m and a history h an agent α and a formula φ gives the chance the agent's choice relative to h is an action resulting in φ, is defined by:* $CoS(m, h, \alpha, \varphi) = \sum_{K \in PosX(m,h,\alpha,\varphi)} B(m, \alpha, K)$.

We can now formulate the central 'responsibility function' definition which will be used to define the truth of the `resp` operator. The responsibility of an agent for an outcome can be seen as the contribution of this agent to the likelihood of this outcome. Formally, the responsibility function is defined as the difference between the chance of success given the current choice of the agent and the minimal chance of success for all the possible choices of the agent.

Definition 11 (Responsibility). *The* responsibility *function* $resp\colon M \times H \times \mathcal{A} \times \mathcal{L} \to C$ *which for a moment m and a history h an agent α and a formula φ gives the responsibility of the agent α in bringing about φ by a choice relative to h is defined by:* $resp(m, h, \alpha, \varphi) = CoS(m, h, \alpha, \varphi) - \min_{h' \in H_m} CoS(m, h', \alpha, \varphi)$.

The truth of formulas are evaluated with respect to moment/history pairs to take into account the dynamic aspect of the actions.

Definition 12. *Relative to a model $\mathcal{M} = \langle M, H, Choices, B, V \rangle$, truth of a formula in a dynamic state $\langle m, h \rangle$, with $m \in h$, are defined as:*

$$\begin{array}{ll}
\langle m,h\rangle \models p & \textit{iff } m \in V(p) \\
\langle m,h\rangle \models \neg\varphi & \textit{iff not } \langle m,h\rangle \models \varphi \\
\langle m,h\rangle \models \varphi \wedge \psi & \textit{iff } \langle m,h\rangle \models \varphi \textit{ and } \langle m,h\rangle \models \psi \\
\langle m,h\rangle \models \Box\varphi & \textit{iff } \forall h' \in H_m, \langle m,h'\rangle \models \varphi \\
\langle m,h\rangle \models X\varphi & \textit{iff } \langle succ(m,h),h\rangle \models \varphi \\
\langle m,h\rangle \models [\alpha\ \texttt{xstit}^{\geq c}]\varphi & \textit{iff } CoS(m,h,\alpha,\varphi) \geq c \\
\langle m,h\rangle \models [\alpha\ \texttt{resp}^{=c}]\varphi & \textit{iff } resp(m,h,\alpha,\varphi) = c
\end{array}$$

We can go back to fig. 1 to evaluate formulas. Relative to moment m_2 and history h_5, the choice made by agent α_1 does not ensure that φ holds, since φ is not true for the second choice of agent α_2 at m_9. But φ is true at m_8 which α_1 believes has a 0.6 chance of happening. Thus we have that $\langle m_2, h_5\rangle \models [\alpha_1\ \texttt{xstit}^{\geq 0.6}]\varphi$. At moment m_2, agent α_1's choice least likely to result in φ is along history h_1 where the chance of success is 0. Thus, relative to moment m_2 and history h_5, the responsibility of α_1 for φ is the difference between the current chance of success of φ (0.6) and the minimal chance of success (0). Therefore, $\langle m_2, h_5\rangle \models [\alpha_1\ \texttt{resp}^{=0.6}]\varphi$.

In case the set C of probabilities is finite, we can express the `resp` operator from the `xstit` operator. We assume moreover that C should be closed under addition and subtraction (modulo 1). It results from this that C is necessarily of the form $C_n = \left\{\frac{k}{n} \mid 0 \leq k \leq n\right\}$ where $n \in \mathbb{N}$. We use the $[\alpha\ \texttt{xstit}^{\geq c}]$ operator as the base operator because it has better properties like additivity and monotonicity but we can define the $[\alpha\ \texttt{xstit}^{=c}]$ operator as:

Definition 13.

$$[\alpha\ \texttt{xstit}^{=\frac{k}{n}}]\varphi \quad := \quad [\alpha\ \texttt{xstit}^{\geq\frac{k}{n}}]\varphi \wedge \neg[\alpha\ \texttt{xstit}^{\geq\frac{k+1}{n}}]\varphi$$

Proposition 1. *The responsibility operator can be expressed in terms of the xstit operator by:*

$$[\alpha\ \texttt{resp}^{=\frac{j-i}{n}}]\varphi \quad \textit{iff}$$

$$[\alpha\ \texttt{xstit}^{\geq\frac{j}{n}}]\varphi \wedge \neg[\alpha\ \texttt{xstit}^{\geq\frac{j+1}{n}}]\varphi \wedge \Box[\alpha\ \texttt{xstit}^{\geq\frac{i}{n}}]\varphi \wedge \neg\Box[\alpha\ \texttt{xstit}^{\geq\frac{i+1}{n}}]\varphi$$

4 Responsibility and degree of causation

Our definition of grades of responsibility can be linked to that of causal contribution defined in [2] by means of the NESS-test. The *NESS-test* says that c is a cause of e iff there is a set of events that is sufficient for e such that: (i) c is a member of the set; (ii) all elements of the set obtain; (iii) c is necessary for the sufficiency of the set. Braham and van Hees define the notion of degree of causation, which can be explained, transcribed in our notation as following.

The *degree of causation* β_i of an agent i is defined as:

$$\beta_i = \frac{|\mathcal{C}_i|}{\sum_{j\in\mathcal{A}} |\mathcal{C}_j|}$$

where $\mathcal{C}_i = \{A \mid A \subseteq \mathcal{A}$ is sufficient for φ and i is φ-critical for $A\}$.

A group of agent A is *sufficient* for φ if and only if $\forall h' \in \mathit{Choice}_G(m, A, h)$, $\langle m, \mathit{succ}(m, h')\rangle \models \varphi$. An agent i is *φ-critical* for A if and only if $\exists i \in \mathcal{A}$ such that $A \setminus i$ is not sufficient for φ.

Braham and van Hees also define a condition for a subjective notion of responsibility in [4]. This notion is not defined with a formal language in that paper, so we give here a formalization of Braham and van Hees's responsibility based on our definitions from the previous section. First, we give an operator for the NESS test.

Definition 14 (NESS operator). *Relative to a model $\mathcal{M}$, truth of the `ness` operator in a dynamic state $\langle m, h\rangle$, with $m \in h$, is defined as:*
$\langle m, h\rangle \models [\alpha\,\mathtt{ness}]\,\varphi$ *iff* $\exists A \subseteq \mathcal{A}$, $\langle m, h\rangle \models [A \cup \{\alpha\}\ \mathtt{xstit}]\varphi \wedge \neg[A\ \mathtt{xstit}]\varphi$
where $\langle m, h\rangle \models [A\,\mathtt{xstit}]\varphi$ *iff* $\forall h' \in \mathit{Choice}_G(m, A, h)$, $\langle \mathit{succ}(m, h'), h'\rangle \models \varphi$

Definition 15 (Non φ-NESS states). *The* non φ-NESS states *function* $R\colon M \times H \times \mathcal{A} \times \mathcal{L} \to \wp(\wp(H))$ *which for a moment m, a history h, an agent α and a formula φ gives the next states in which $\alpha$is not NESS for φ, given the agent's current choice determined by h:* $R(m, h, \alpha, \varphi) =$
$\{K \in \mathit{States}(m, \alpha) \mid \forall h' \in K \cap \mathit{Choice}(m, \alpha, h),\ \langle m, h'\rangle \models \neg\,[\alpha\,\mathtt{ness}]\,\varphi\}$.

Definition 16 (Avoidance potential). *The* avoidance potential *function* $\rho\colon M \times H \times \mathcal{A} \times \mathcal{L} \to C$ *which for a moment m and a history h an agent α and a formula φ gives the chance the agent's choice relative to h is an action avoiding φ, is defined by:* $\rho(m, h, \alpha, \varphi) = \sum_{K \in R(m,h,\alpha,\varphi)} B(m, \alpha, K)$.

Definition 17 (Responsibility operator). *Relative to a model $\mathcal{M}$, truth of the `resp_BvH` operator in a dynamic state $\langle m, h\rangle$, with $m \in h$, is defined as:* $\langle m, h\rangle \models [\alpha\,\mathtt{resp\_BvH}]\,\varphi$ *iff (i)* $\langle m', h\rangle \models [\alpha\,\mathtt{ness}]\,\varphi$ *and (ii)* $\exists h' \in H_{m'}$, *such that* $\rho(m', h', \alpha, \varphi) > \rho(m', h, \alpha, \varphi)$ *where* $m' = \mathit{prec}(m, h)$.

5 Examples

We are going to compare the results of responsibility operator with those from Braham and van Hees in [2], using examples that illustrate a variety of situations and represent some modeling problems such as overdetermination or the knowledge of the agents. Even though Braham and van Hees's degree of causation is not probabilistic, we want to show that we also get correct results with our operator in non probabilistic settings when considering uniform probabilities.

Example 1 (Toxins [3]). Firms 1, 2 and 3 dumped different toxins in a river, denoted by T_1, T_2, and T_3 respectively. The combination of the three actions was necessary to kill all the fish in the river. Using our framework, the responsibility of each firm is therefore $\frac{1}{4}$ and is in accordance with the intuition that each agent should be equally responsible. The result obtained using the degree of causation gives $\beta_1 = \beta_2 = \beta_3 = \frac{1}{3}$ which also agree with the intuition. The value is not the same as the one we have with our framework, but the relative responsibility among the agents is the same. For simple situations like the toxin example our framework gives similar results to that of the degree of causation but it is not always the case for more complex situations.

Now, suppose that Firm 3 transfers its activities to Firm 1 so that Firm 3 ceases to dump T_3 in the river. Firm 1 now dumps toxins T_1 and T_3 into the river and Firm 2 dumps T_2. In this situation, it seems intuitive to say that Firm 1 takes over Firm 3 responsibility along with the activity and that Firm 1 is now twice as responsible as Firm 2. In our framework, the responsibility of Firm 1 is $\frac{1}{2}$ and the responsibility of Firm 2 is $\frac{1}{4}$ and agrees with the intuition. However, the degree of causation defined previously ascribes an equal responsibility to both firms. To circumvent this problem, Braham and van Hees have to extend the notion of degree of causation to complex actions and get the expected result: $\beta_1^* = \frac{2}{3}$ and $\beta_2^* = \frac{1}{3}$ with gives the same relative values as our framework.

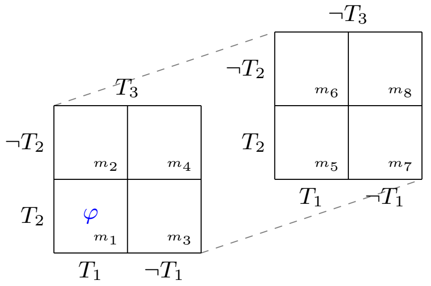


Fig. 2: Three firms dumping toxins

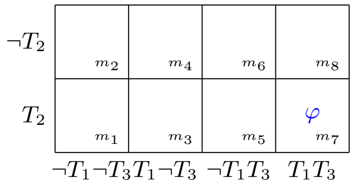


Fig. 3: Two firms dumping toxins

Example 2 (Vote with over-determination [7]). Six voters $\{a, b, c, d, e, f\}$ have to vote in favor or against a proposal. The minimal winning coalitions in favor of the proposal are $\{a, b, c\}$, $\{a, b, d\}$ and $\{a, e, f\}$. The outcome of the vote was that $S = \{a, b, c, d, e\}$ voted in favor. This is a case of over-determination as either $\{a, b, c\}$ or $\{a, b, d\}$ was sufficient to ensure the proposal was accepted. It seems intuitive to say that each of a, b, c and d undoubtedly bears some responsibility for the outcome, but what about e? e is not a cause of the proposition being accepted; however, e's vote *could* have contributed to the proposition being accepted (had f voted in favor) and should therefore bear some part of the responsibility. Both the degree of causation and Braham and van Hees's responsibility operator consider that e is not responsible for the outcome of the vote because it is not a cause of the outcome. In our framework, a has a responsibility of $\frac{17}{32}$, b of $\frac{9}{32}$ and c, d have an equal responsibility of $\frac{3}{32}$ and e has a responsibility of $\frac{5}{32}$. f has no responsibility for the outcome because he voted against the proposition. In this situation, the results we get with our framework conform to our expectations whereas neither the degree of causation nor Braham and van Hees's definition of responsibility do.

Example 3 (Evening out). John and Mary want to go out together in the evening. They can either go to a restaurant, to the cinema or to a concert. They have agreed to go to a restaurant and John is certain that Mary is going to keep her promise. If John chooses to do something else, then it is reasonable to say that he is responsible for them not meeting for their evening out. Assuming that John believes that Mary has 80% chances of going to the agreed place and 10% chances to go to each of the others, according to our framework, John has a responsibility of 0.7 if he chooses not to go to the restaurant. If Mary has the same beliefs about John, but goes to the restaurant, her responsibility is 0. The degree of causation is the same of Mary and John, which does not represent accurately the responsibility in this situation where knowledge is involved. Braham and van Hees's responsibility says that John is responsible and Mary isn't, but doesn't give a grade of responsibility for each agent.

6 Conclusion

This paper extends the probabilistic XSTIT logic by adding a responsibility operator. This operator assign to each agent a level of responsibility depending on the subjective beliefs of this agent regarding the actions of the other agents. We have also shown that our responsibility operator performs as well or better than the degree of causation and the responsibility defined by Braham and van Hees. There is opportunity for further research in different directions: first, the framework supposes that each agent is aware of all his possible choices, which is not always realistic; secondly, it is assumed that the agents always know all the consequences of each outcome, which is not always the case; and finally, it would be interesting to investigate the deontic aspect of responsibility.

References

1. N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world.* Oxford University Press, 2001.
2. Matthew Braham and Martin van Hees. Degrees of causation. *Erkenntnis*, 71:323 – 344, 2009.
3. Matthew Braham and Martin van Hees. An anatomy of moral responsibility. *Mind*, 121(483):601–634, 2012.
4. Matthew Braham and Martin van Hees. Voids or fragmentation: Moral responsibility for collective outcomes. 2014.
5. Jan Broersen. Probabilistic stit logic and its decomposition. *International Journal of Approximate Reasoning*, (54):467477, 2013.
6. Jan Broersen, Andreas Herzig, and Nicolas Troquard. From coalition logic to STIT. *Electronic Notes in Theoretical Computer Science*, 157(4):23–35, 2006. Proceedings of Logic and Communication in Multi-Agent Systems (LCMAS 2005).
7. Dan S Felsenthal and Moshé Machover. A note on measuring voters' responsibility. *Homo Oeconomicus*, 26(2):259–271, 2009.

8. Peter Vallentyne. Brute luck and responsibility. *Politics, Philosophy & Economics*, 7(1):57–80, 2008.

Modularising and Promoting Interoperability for Event-B Specifications using Institution Theory

Marie Farrell, Rosemary Monahan, and James F. Power

Maynooth University, Maynooth, Co. Kildare, Ireland
mfarrell@cs.nuim.ie

Abstract. We motivate the need for modularisation constructs in the Event-B formal specification language. We utilise the specification building operators of the theory of institutions to provide these modularisation constructs and we present an example of a traffic-light simulation to illustrate our approach.

Keywords: Event-B; institutions; refinement; formal methods; modular specification

1 Introduction and Motivation

Modern software development focuses on model-driven engineering: the construction, maintenance and integration of software models, ranging from high-level design documents (often expressed through diagrams) down to program code. For example, a model-based approach to developing software might start with the construction of a design model, such as a UML class diagram, developing class functionality via state machines, and then forward engineering this to program code.

In the field of formal software development we can prove the correctness of a particular piece of software by reasoning logically about the system. Just as for non-formal development, it can be beneficial to model the aspects of a system using a variety of specialised formalisms to ensure different aspects of its correctness. In formal software engineering we can map between these levels of abstraction in a verifiable way through a process known as refinement, which can take place within a single modelling language, or between languages at different levels of abstraction [10]. The ideal scenario has been described as a "theory supermarket", in which a developer can shop for suitable theories with confidence that they will work together [5].

This paper is centered around an illustrative example of a specification in Event-B, inspired by one in the *Rodin User's Handbook* [8]. In Section 2 we provide an overview of the Event-B formalism, identify its limitations and discuss related work. We have identified the theory of institutions as a means of enhancing the Event-B formalism and we define an institution for Event-B in Section 3. The definition of $\mathcal{EVT}$, our institution for Event-B, enables us to utilise the specification-building operators provided by institutions and to re-cast our example in modular form. We address refinement in Section 4 since this is of central importance in Event-B, and show how this too can be modularised using institutional specification building operations. We summarise our contributions and outline future directions in Section 5.

2 Background: specification and refinement in Event-B

Many tools and formalisms have been developed to facilitate the process of software specification, refinement and verification. Event-B is an industrial-strength language for system-level modelling and verification that combines an event-based logic with basic set theory [1]. A key feature of Event-B is its support for formal refinement, which allows a developer to write an abstract specification of a system and then to gradually add complexity in a provable correct way [11]. The *Rodin Platform*, an integrated development environment for Event-B, ensures the safety of system specifications and refinement steps by generating appropriate proof-obligations, and then discharging these via support for various theorem provers [8]. Event-B has been used extensively in a number industrial projects, such as the Paris Métro Line 14, and is a relatively mature language.

2.1 Modelling the Traffic Lights

We have provided an illustrative example of an Event-B model of a traffic lights system that is inspired by one in the *Rodin User's Handbook* [8]. Figure 1 presents an Event-B machine for a traffic lights system with one light signalling cars and one light signalling pedestrians. In general, machine specifications model the dynamic behaviour of a system and can contain variable declarations (lines 2-3), invariants (lines 4-7) and event specifications (lines 8-33). The goal of the specification is to ensure that it is never the case that both cars and pedestrians receive the "go" signal at the same time (represented by boolean flags on line 3).

Figure 1 specifies five different events (including a starting event called `Initialisation` (lines 9-13)). An event is composed of a guard (predicate) and an action which is represented as a before-after predicate relating the new values of the variables to the old. Events can happen in any order once their guards evaluate to true and the theorem provers check that each invariant is not violated by any event. For example, the `set_peds_go` event as specified on lines 14-19, has one guard expressed as a boolean expression (line 16), and one action, expressed as an assignment statement (line 18). In general an event can contain many guards and actions, though a variable can only be assigned to once (and assignments occur in parallel) [8].

In addition to machine specifications, *contexts* in Event-B can be used to model the static properties of a system (constants, axioms and carrier sets). Figure 2 provides a context giving a specification for the data-type $COLOURS$ and uses the axiom on line 7 to explicitly restrict the set to only contain the constants *red*, *green* and *orange*.

```
MACHINE mac1
VARIABLES
  cars_go, peds_go
INVARIANTS
   inv1 : cars_go ∈ BOOL
   inv2 : peds_go ∈ BOOL
   inv3 : ¬ (peds_go = true ∧ cars_go = true)
EVENTS
  Initialisation
    begin
       act1 : cars_go := false
       act2 : peds_go := false
    end
  Event set_peds_go ≙
    when
       grd1 : cars_go = false
    then
       act1 : peds_go := true
    end
  Event set_peds_stop ≙
    begin
       act1 : peds_go := false
    end
  Event set_cars_go ≙
    when
       grd1 : peds_go = false
    then
       act1 : cars_go := true
    end
  Event set_cars_stop ≙
    begin
       act1 : cars_go := false
    end
END
```

Fig. 1: Event-B machine specification for a traffic system, with cars and pedestrians controlled by boolean flags.

```
CONTEXT ctx1
SETS
  COLOURS
CONSTANTS
  red, green, orange
AXIOMS
   axm1 : partition(COLOURS, {red}, {green}, {orange})
END
```

Fig. 2: Event-B context specification for the colours of a set of traffic lights.

```
MACHINE mac2
refines mac1
SEES ctx1
VARIABLES
  cars_colour, peds_colour, buttonpushed
INVARIANTS
   inv1 : peds_colour ∈ {red, green}
   inv2 : (peds_go = TRUE) ⇔ (peds_colour = green)
   inv3 : cars_colour ∈ {red, green}
   inv4 : (cars_go = TRUE)⇔ (cars_colour = green)
   inv5 : buttonpushed ∈ BOOL
EVENTS
  Initialisation
    begin
       act1 : cars_colour := red
       act2 : peds_colour := red
    end
  Event set_peds_green ≙
  refines set_peds_go
    when
       grd1 : cars_colour = red
       grd2 : buttonpushed = true
    then
       act1 : peds_colour := green
       act2 : buttonpushed := false
    end
  Event set_peds_red ≙
  refines set_peds_stop
    begin
       act1 : peds_colour := red
    end
  Event set_cars_green ≙
  refines set_cars_go
    when
       grd1 : peds_colour = red
    then
       act1 : cars_colour := green
    end
  Event set_cars_red ≙
  refines set_cars_stop
    begin
       act1 : cars_colour := red
    end
  Event press_button ≙
    begin
       act1 : buttonpushed := true
    end
END
```

Fig. 3: A refined Event-B machine specification for a traffic system, with cars and pedestrians controlled by a button-activated pedestrian light.

Figure 3 shows an Event-B machine specification, `mac2`, which refines `mac1` from Figure 1. `mac2` refines `mac1` by first introducing the new context on line 3 and then by replacing the truth values used in the abstract machine with new values from the carrier set $COLOURS$. During refinement, the user typically supplies a *gluing invariant* relating properties of the abstract machine to their counterparts in the concrete machine [8]. The gluing invariants shown in lines 8 and 10 of Figure 3 define a one-to-one mapping between the concrete variables introduced in mac2 and the abstract variables of mac1. As specified in lines 7 and 9, the new variables (*peds_colour* and *cars_colour*) can be either *red* or *green*, thus the gluing invariants map *true* to *green* and $false$ to *red*.

Event-B permits the addition of new variables and events - *buttonpushed* on line 5 and `press_button` on lines 44-47. Also, the existing events from `mac1` are renamed to reflect refinement; for example, on lines 18-19 the event `set_peds _green` is declared to refine `set_peds_go`. This event has also been altered via the addition of a guard (line 22) and an action (line 25) which incorporate the functionality of a button-controlled pedestrian light.

2.2 Limitations of Event-B

Although a very mature formalism, we believe there are two main areas where the Event-B language needs further improvement:

Modularity: The given example highlights features of the Event-B language, but notice how, in Figure 1 the same specification has to be provided twice. The events `set_peds_go` and `set_peds_stop` are equivalent, modulo renaming of variables, to `set_cars_go` and `set_cars_stop`. Ideally, writing and proving the specification for these events should only happen once. Therefore, we can identify one weakness of Event-B as its lack of well-developed modularisation constructs and it is not easy to combine specifications in Event-B with those written in other formalisms [7].

Interoperability: Large software systems are often at such a level of complexity that no single formalisation, encoding or abstraction of can aptly represent and reason about the whole system. This results in the system being modelled numerous times, often in separate formalisms, thus requiring proof repetition. For example, when developing software using Event-B, it is at least necessary to transform the final concrete specification into a different language to get an executable implementation.

2.3 Related Work

One suggested method of providing modularity for Event-B specifications is model decomposition, originally proposed by Abrial (shared variable [2]) and Butler (shared event [15]), and later developed as a plugin for the *Rodin Platform* [16]. The *shared variable* approach partitions the model into sub-models based on events sharing the same variables. The *shared event* method partitions the model based on variables participating in the same events. This approach is quite restrictive in that it is not possible to refer to the same element across sub-models. Also, it is impossible to select which invariants are allocated to each sub-model, currently, only those relating to variables of the sub-model are included.

Another approach is the modularisation plugin for *Rodin* [7], which is based on the *shared variable* method outlined above. Here, *modules* split up an Event-B model and are paired with an *interface* describing conditions for incorporating the module into another. Module interfaces list the operations contained in the module. Modules are similar to machines but they may not specify events. The events of a machine which imports an interface can see the visible constants, sets and axioms, call the imported operations, and the interface variables and invariants are added to the machine. Although similar to the *shared variable* approach proposed by [2] this method is less restrictive, as invariants can be included in the module interface.

Both of these *Rodin* plugins provide some degree of modularisation for Event-B but they do not directly enhance the Event-B formalism itself nor do they provide scope for the interoperability of Event-B with other formalisms and/or logics.

Current approaches to interoperability in Event-B consist of a range of *Rodin* plugins to translate to/from Event-B but these lack a solid logical foundation. Examples include UML-B [17] and EventB2JML [3].

In summary, the existing approaches to addressing modularity and interoperability issues in Event-B tend to be somewhat ad hoc, causing difficulties for interaction, proof sharing and maintainability. The goal of our research is to develop a set of modularisation constructs for Event-B that will be sufficiently generic, so that they are well understood (particularly in formal terms), and so that they can map easily to similar constructs in other formalisms. We also intend to provide scope for the interoperability of Event-B with other formalisms as part of our solution.

The core to ensuring modularisation and interoperability in model-driven engineering is *meta-modelling*: the modelling of modelling languages. Similarly, when dealing with logic-based formalisms that include specification, refinement and proof, the key to ensuring interoperability is a suitable meta-logical framework, that will allow for the specification of specification languages. These ideas have a long history in logic, going back at least to Tarski's work in the 1930s on the definition and classification of consequence relations [18].

3 Institutions - a Meta-Logical Framework

Originating from Tarski's work on metamathematics and consequence, the theory of institutions provides a meta-logical framework in which a set of specification building operators can be defined allowing you to write, modularise and build up specifications in a formalism-independent manner [6, 18]. In order to represent a formalism/logic in this way, the syntax and semantics for it must first be defined in a uniform way using some basic constructs from category theory [14]. Institutions have been defined for many logics and formalisms, including programming-related formal languages such as UML and CSP [9, 13].

A huge benefit of this approach is that it facilitates the use of specification building operators that provide modularisation constructs to any logic/formalism presented in this way. Examples of formalisms that have been improved by using institutions are those for UML state machines [9] and CSP[13]. Readers familiar with Unifying Theories of Programming may note that the notion of institutions in this way is similar to the notion of a "theory supermarket" in which one can shop for theories with confidence that they will work together [5].

Once a formalism/logic has been described using institutions a range of specification building operators become available [14]. These operators facilitate the modularisation of specifications and describe how specifications can be combined in different formalisms/logics. They facilitate the combination (and, $+$, $\cup$), extension (then), hiding (hide via, reveal) and renaming (with) of specifications.

We have represented Event-B as a logic in the theory of institutions, as such, we gain the use of specification building operators to increase modularity and an embedding in a framework designed to promote and facilitate interoperability with other formalisms. Since a key feature of Event-B is refinement, it is vital that any representation of Event-B maintain the same notion of refinement. The theory of institutions already accounts for this so there is no need to redefine it [14]. Another major benefit of representing Event-B in terms of institutions is that it provides a formal semantics for Event-B that is fully rooted in a mathematical foundation.

3.1 Defining $\mathcal{EVT}$, an institution for Event-B

$\mathcal{EVT}$, our formalisation of Event-B in terms of institutions is based on splitting an Event-B specification into two parts:

- A data part, which can be defined using some standard institution such as that for algebra or first-order logic. We have chosen $\mathcal{FOPEQ}$, the institution for first order predicate logic with equality [14], since it most closely matches the kind of data specification needed.
- An event part, which defines a set of events in terms of formula constraining their before- and after- states. Our specification here is closely based on $\mathcal{UML}$, an institution for UML state machines [9].

While we do not have space to present the details fully formally here, they are not more complex than those normally used for first-order logic, with appropriate assignments for the free variables named in the event specification variables.

In order to build an institution for Event-B, which we call $\mathcal{EVT}$, it is necessary to specify and verify a series of definitions (using category theory) for its syntax and semantics. Once these language elements have been specified, the next step is to verify that the resulting metalogical structure is actually a valid institution. This is ensured by proving the *satisfaction condition* which states, in formal terms, the basic maxim of institutions, that "truth is invariant under change of notation". We can only outline this process here, but full details of the institution $\mathcal{EVT}$ and the associated proofs are available from our website.[1]

I. Signatures. A signature over $\mathcal{EVT}$ describes the vocabulary that we are allowed to use when writing Event-B specifications, and consists of names for sorts, operations, predicates, events and variables. We assume that each operation, predicate and variable name is appropriately indexed by its sort and arity. Signature *morphisms* provide a mechanism for moving between vocabularies while repspecting arities, sort-indexing and initialisation events. By construction, these morphisms can be extended in a uniform way to models and sentences.

II. Models. A *data state* consists of a set of values for each of the variables in the signature corresponding to the declared sort of the variable. A possible execution of a machine is then represented by a *trace*, which is just a sequence of data states, with each step in this sequence being labelled by an event name. Finally, a *model* of an $\mathcal{EVT}$ signature is a set of such traces, specified as a relation over the states whose constituent tuples are labelled by event names.

III. Sentences. A sentence over $\mathcal{EVT}$ is then an element of an Event-B specification written using the names from the signature. In the *Rodin Platform* Event-B sentences are presented (with suitable syntactic sugaring) as:

```
$I(\overline{x})$
Event e $\widehat{=}$
  when
    guard-name : $G(\overline{x})$
  then
    act-name : $A(\overline{x}, \overline{x}')$
  end
```

where $I(\overline{x})$ and $G(\overline{x})$ are predicates representing the invariant(s) and guard(s) respectively over the set of variables $\overline{x}$. In Event-B, actions are interpreted as before-after predicates i.e. the statement $x := x+1$ is interpreted as the predicate $x' = x + 1$. Therefore, a predicate of the form $A(\overline{x}, \overline{x}')$ represents the action(s) over the sets of variables $\overline{x}$ and $\overline{x}'$. Here $\overline{x}'$ is the same set of variables as $\overline{x}$ but with all of the names primed.

Based on this we can define the syntax of $\mathcal{EVT}$ in terms of two types of sentence.

- The first kind of sentence is an *invariant definition*, which is simply a predicate $\phi(\overline{x})$ over the variables in the signature.
- The second kind of sentence represents an *event definition* and consists of a pairing $e \mathrel{\widehat{=}} \psi(\overline{x}, \overline{x}')$ where e is an event name and $\psi(\overline{x}, \overline{x}')$ is a $\mathcal{FOPEQ}$ predicate corresponding to $G(\overline{x}) \wedge A(\overline{x}, \overline{x}')$ in the above Event-B sentence.

IV. Satisfaction. The satisfaction of $\mathcal{EVT}$-sentences by $\mathcal{EVT}$-models is split into satisfaction for each kind of sentence.

[1] http://www.cs.nuim.ie/∼mfarrell/

```
 1 spec TwoBools over 𝓕𝓞𝓟𝓔𝓠
 2   Bool
 3   then
 4     ops
 5       i_go, u_go : Bool
 6     preds
 7       ¬ (i_go = true ∧ u_go = true)


 8 spec LightAbstract over 𝓔𝓥𝓣
 9   TwoBools
10   then
11     Initialisation
12       begin
13         act1  : i_go := false
14       end
15     Event set_go ≙
16       when
17         grd1 : u_go = false
18       then
19         act1 : i_go := true
20       end
21     Event set_stop ≙
22       then
23         act1 : i_go := false
24       end

25 spec mac1 over 𝓔𝓥𝓣
26   (LightAbstract with σ1)
27     and (LightAbstract with σ2)
28
29   where
30     σ1 = {i_go ↦ cars_go,
31           u_go ↦ peds_go,
32           set_go ↦ set_cars_go,
33           set_stop ↦ set_cars_stop}
34
35     σ2 = {i_go ↦ peds_go,
36           u_go ↦ cars_go,
37           set_go ↦ set_peds_go,
38           set_stop ↦ set_peds_stop}
```

Fig. 4: A modular institution-based presentation corresponding to the abstract machine `mac1` in Fig 1.

Satisfaction of invariant sentences: If some predicate $\phi(\overline{x})$ is given as an invariant, then an $\mathcal{EVT}$-model m satisfies $\phi(\overline{x})$ if that formula evaluates to true in each data state of the model.

Satisfaction of event sentences: Given an definition of an event e by some predicate $\psi(\overline{x}, \overline{x}')$, then an $\mathcal{EVT}$-model m satisfies this sentence if the predicate $\psi(\overline{x}, \overline{x}')$ evaluates to true for every pair of states in the model labelled by e.

In order to ensure that an institution $\mathcal{EVT}$ has good modularity properties it is necessary to carry out some category theoretic proofs. In particular, pushouts must exist in the category of signatures and the institution must have the amalgamation property [14].

3.2 An Example of specification-building in $\mathcal{EVT}$

Defining $\mathcal{EVT}$, an institution for Event-B, allows us to restructure Event-B specifications using the standard specification building operators for institutions [14]. Thus $\mathcal{EVT}$ provides a means for writing down and splitting up the components of an Event-B system, facilitating increased modularity for Event-B specifications. Figure 4 is a presentation (set of sentences) over the institution $\mathcal{EVT}$ corresponding to the Event-B machine `mac1` defined in Figure 1. The presentation in Figure 4 consists of three specifications:

Lines 1-7: The specification TwoBools, technically in $\mathcal{EVT}$, can be presented as a pure specification over $\mathcal{FOPEQ}$, declaring two boolean variables constrained to have different values. The predicate on line 7 here corresponds to the invariant on line 7 of Figure 1.

Lines 8-24: LightAbstract is a specification over $\mathcal{EVT}$ for a single traffic light that extends TwoBools (then). It contains the events `set_go` and `set_stop`, with a constraint that a light can only be set to "go" if its opposite light is not.

Lines 25-38: The specification mac1 combines (and) two versions of LightAbstract each with a different signature morphism (σ_1 and σ_2) mapping the specification variables and event names to those in the Event-B machine.

Notice that the specification for each individual light had to be explicitly written down twice in the Event-B machine in Fig 1. In our modular institution-based presentation it is only necessary to have one light specification and simply supply the required variable and event mappings. In this way, $\mathcal{EVT}$ adds much more modularity than is currently present in Event-B, and these constructs are well defined in the theory of institutions providing a formal mathematical foundation for modularisation in Event-B.

4 Refinement in the $\mathcal{EVT}$ institution

Event-B supports three forms of machine refinement: the refinement of event internals (guards and actions) and invariants; the addition of new events; and the decomposition of an event into several events [4, 8]. It is therefore essential that any formalisation of Event-B be capable of capturing these concepts. The theory of institutions provides support for all three types of Event-B refinement as it is, in fact, equipped with a well-defined notion of refinement [14].

4.1 A modular, refined specification

Figure 5 contains a presentation over $\mathcal{EVT}$ corresponding to the main elements of the Event-B specification mac2 presented in Figures 2 and 3. Here, we present three data specifications over $\mathcal{FOPEQ}$ and three event specifications over $\mathcal{EVT}$.

Lines 1-11: We specify the *Colours* data type with a standard data specification, as can be seen in Figure 2. The specification TwoColours describes two variables of type *Colours* constrained not both be green at the same time. This corresponds to the gluing invariants on lines 8 and 10 of Figure 3. The specification modularisation constructs used in Figure 5, allow these properties to be handled distinctly and in a manner that facilitates comparison with the TwoBools specification on lines 1-7 of Figure 4.

Lines 12-28: A specification for a single light is provided in LightRefined which uses TwoColours to describe the colour of the lights. As was the case with LightAbstract in Figure 4, the specification makes clear how a single light operates. An added benefit here is that a direct comparison with the abstract specification can be done on a per-light basis.

Lines 29-46: The specifications BoolButton and ButtonSpec account for the part of the mac2 specification that requires a button. These details were

```
 1  spec Colours over FOPEQ
 2    then
 3      sorts
 4      Colours free with red|green|orange

 5  spec TwoColours over FOPEQ
 6      Colours
 7      then
 8        ops
 9          icol, ucol : Colours
10        preds
11          ¬(icol = green ∧ ucol = green)

12  spec LightRefined over EVT
13    TwoColours
14    then
15      Initialisation
16        begin
17           act1 : icol := red
18        end
19      Event set_green ≙
20        when
21           grd1 : ucol = red
22        then
23           act1 : icol := green
24        end
25      Event set_red ≙
26        then
27           act1 : icol := red
28        end

29  spec BoolButton over FOPEQ
30    Bool
31    then
32      ops
33        button : Bool

34  spec ButtonSpec over EVT
35    BoolButton
36    then
37      Event gobutton ≙
38        when
39           grd1 : button = true
40        then
41           act1 : button := false
42        end
43      Event pushbutton ≙
44        then
45           act1 : button := true
46        end

47  spec mac2 over EVT
48    (LightRefined  with σ3)
49    and
50    (LightRefined and
51          (ButtonSpec with σ5)
52          with σ4)
53
54  where
55    σ3 = {i_col ↦ cars_colour,
56          u_col ↦ peds_colour,
57          set_green ↦ set_cars_green,
58          set_red ↦ set_cars_red}
59
60    σ4 = {i_col ↦ peds_colour,
61          u_col ↦ cars_colour,
62          set_green ↦ set_peds_green,
63          set_red ↦ set_peds_red}
64
65    σ5 = {gobutton ↦ press_button}
```

Fig. 5: A modular institution-based presentation corresponding to the refined machine `mac2` specified in Fig 3.

woven through the code in Figure 3 (lines 5, 11, 22, 25, 45) but the specification-building operators allow us to modularise the specification and group these related definitions together, clarifying how the button actually operates.

Lines 47-65: Finally, to tie this all together we must combine a copy of LightRefined with a specification corresponding to the sum (and) of LightRefined and ButtonSpec with appropriate signature morphisms. This second specification combines the event `gobutton` in ButtonSpec with the event `set_green` in LightRefined thus accounting for `set_peds_green` in Figure 3. One small issue involves making sure that the name replacements are done correctly, and in the correct order, hence the bracketing on lines 48-52 is important.

The combination of these specifications involves merging two events with different names: `gobutton` from ButtonSpec with the event `set_green` from LightRefined. To ensure that these differently-named events are combined into an event of the same name we use the signature morphism σ_5 to give `gobutton` the same name as `set_green` before combining them. By ensuring that the events have the same name, and combines both events' guards and actions and the morphism σ_4 names the resulting event `set_peds_green`. The resulting specification will also contain the event `pushbutton`.

Note that the labels given to guards/actions are syntactic sugar to make the specification aesthetically resemble the usual Event-B notation for guards/actions.

5 Conclusion and Future Work

Our specification of $\mathcal{EVT}$ has enabled us to address the limitations in the Event-B language that we have identified in Section 2.2 as follows:

Modularity: By defining $\mathcal{EVT}$ and carrying out the appropriate proofs, we gain access to an array of generic specification building operators [14]. These facilitate the combination (and, +, ∪), extension (then), hiding (hide via, reveal) and renaming via signature morphism (with) of specifications. Representing Event-B in this way provides us with a mechanism for combining and parameterising specifications. Most importantly, these constructs are formally defined, a crucial issue for a language used in formal modelling.

Interoperability: Institution comorphisms can be defined enabling us to move between different institutions, thus providing a mechanism by which a specification written over one institution can be represented as a specification over another. Devising meaningful institutions and corresponding morphisms to/from Event-B provides a mechanism for not only ensuring the safety of a particular specification but also, via morphisms, a platform for integration with other formalisms and logics.

Another benefit of developing an institution-based specification for Event-B is that it provides a formal semantics for the language, something that has not been explicitly developed thus far, although developed informally [1].

We have successfully specified an institution for the Event-B formalism and proved the relevant properties that allow for the use of the modularisation constructs. Our current task is that of implementation using the Heterogeneous Tool-Set, Hets, a framework for institution-based heterogeneous specifications [12]. A significant future challenge is the integration of proofs for Event-B, developed using the *Rodin* platform, into the more general Hets environment.

Devising meaningful institutions and corresponding morphisms to/from Event-B provides a mechanism for not only ensuring the safety of a particular specification but also, via morphisms, a potential for integration with other formalisms. Interoperability and heterogeneity are significant goals in the field of software engineering, and we believe that the work presented in this paper provides a basis for the integration of Event-B with other formalisms based on the theory of institutions.

References

1. J.-R. Abrial. *Modeling in Event-B: System and Software Engineering.* Cambridge University Press, New York, NY, USA, 1st edition, 2010.

2. J.-R. Abrial and S. Hallerstede. Refinement, decomposition, and instantiation of discrete models: Application to Event-B. *Fundamenta Informaticae*, 77(1-2):1–28, 2007.
3. N. Cataño, T. Wahls, C. Rueda, V. Rivera, and D. Yu. Translating B machines to JML specifications. In *27th Annual ACM Symposium on Applied Computing*, pages 1271–1277, New York, NY, USA, 2012. ACM.
4. K. Damchoom, M. Butler, and J.-R. Abrial. Modelling and proof of a tree-structured file system in Event-B and Rodin. In *Formal Methods and Software Engineering*, volume 5256 of *LNCS*, pages 25–44. 2008.
5. J. Fitzgerald, P. G. Larsen, and J. Woodcock. Foundations for model-based engineering of systems of systems. In *Complex Systems Design & Management*, pages 1–19. Springer, 2014.
6. J. A. Goguen and R. M. Burstall. Institutions: abstract model theory for specification and programming. *Journal of the ACM*, 39(1):95–146, 1992.
7. A. Iliasov, E. Troubitsyna, L. Laibinis, A. Romanovsky, K. Varpaaniemi, D. Ilic, and T. Latvala. Supporting reuse in Event-B development: Modularisation approach. In *Abstract State Machines, Alloy, B and Z*, volume 5977 of *LNCS*, pages 174–188. 2010.
8. M. Jastram and P. M. Butler. *Rodin User's Handbook: Covers Rodin V.2.8.* CreateSpace Independent Publishing Platform, USA, 2014.
9. A. Knapp, T. Mossakowski, M. Roggenbach, and M. Glauer. An institution for simple UML state machines. In *Fundamental Approaches to Software Engineering*, volume 9033 of *LNCS*, pages 3–18. 2015.
10. C. Morgan. *Programming from Specifications.* Prentice Hall, U.K., 2nd edition, 1998.
11. C. Morgan, K. Robinson, and P. Gardiner. *On the Refinement Calculus.* Springer, 1988.
12. T. Mossakowski, C. Maeder, and K. Lüttich. The heterogeneous tool set, Hets. In *Tools and Algorithms for the Construction and Analysis of Systems*, volume 4424 of *LNCS*, pages 519–522, 2007.
13. T. Mossakowski and M. Roggenbach. Structured CSP - a process algebra as an institution. In *Recent Trends in Algebraic Development Techniques*, volume 4409 of *LNCS*, pages 92–110. 2007.
14. D. Sanella and A. Tarlecki. *Foundations of Algebraic Specification and Formal Software Development.* Springer, 2012.
15. R. Silva and M. Butler. Shared Event Composition/Decomposition in Event-B. In *International Symposium on Formal Methods for Components and Objects*, pages 122–141. Springer, 2010.
16. R. Silva, C. Pascal, T. S. Hoang, and M. Butler. Decomposition tool for Event-B. *Software: Practice and Experience*, 41(2):199–208, 2011.
17. C. Snook and M. Butler. UML-B: Formal modeling and design aided by UML. *ACM Trans. on Software Engineering and Methodology*, 15(1):92–122, 2006.
18. A. Tarski. On some Fundamental Concepts of Metamathematics. In *Logic, semantics, metamathematics: papers from 1923 to 1938*, chapter 3. Hackett Publishing, 1983.

Short Elementary Cuts in Countable Models of Compositional Arithmetical Truth

Michał Tomasz Godziszewski

Department of Logic, Institute of Philosophy, University of Warsaw
ul. Krakowskie Przedmieście 3, 00-047 Warsaw, Poland
mtgodziszewski@gmail.com

1 Introduction

We study certain model-theoretic properties of countable recursively saturated models of arithmetic. The study of possible semantics for arithmetized languages in nonstandard models has been a lively research field since the seminal paper of A. Robinson [13]. Our primary inspiration for examining mathematical features of such structures, and recursively saturated in particular, is that every countable recursively saturated model of Peano Arithmetic supports a great variety of nonstandard satisfaction classes that can serve as models for formal theories of truth - those models allow to investigate the role of arithmetic induction in semantic considerations. In the other direction, nonstandard satisfaction classes are used as a tool in model theoretic constructions providing answers to questions in the model theory of formal arithmetic and often make it possible to solve problems that do not explicitly involve nonstandard semantics.

A satisfaction class is a subset of the model of PA corresponding to the notion of truth in nonstandard models (see S. Krajewski [11], F. Engstrom [4] and H. Kotlarski [10]). We provide the definition via a formal truth theory axiomatizing Tarski's compositional conditions (with an undisturbing abuse of terminology, substituting truth for satisfiability) - Tarski's theorem on undefinability of truth informs us that it is impossible to obtain a faithful truth predicate for a formal theory within its object language, thus one of the approaches to investigation of the concept of truth is axiomatizing it with the use of a fresh unary predicate $Tr(x)$ with the intended meaning that *the sentence with the Gödel number x is true.*

Definition 1 (Stratified Compositional Truth Theory)
Stratified Compositional Truth Theory (CT^-) is an axiomatic theory obtained from PA by adjoning to it the following axioms:
(CT1) $\forall s,t \in Trm[Tr(s=t) \equiv val(s)=val(t)]$,
(CT2) $\forall x[Sent_{\mathcal{L}}(x) \Rightarrow (Tr(\neg x) \equiv \neg Tr(x))]$,
(CT3) $\forall x,y[(Sent_{\mathcal{L}}(x \wedge y)) \Rightarrow (Tr(x \wedge y) \equiv Tr(x) \wedge Tr(y))]$,
(CT4) $\forall x,y[(Sent_{\mathcal{L}}(x \vee y)) \Rightarrow (Tr(x \vee y) \equiv Tr(x) \vee Tr(y))]$,
(CT5) $\forall v,x[Sent_{\mathcal{L}}(\forall vx) \Rightarrow (Tr(\forall vx) \equiv \forall t Tr(x(t/v)))]$,
(CT6) $\forall v,x[Sent_{\mathcal{L}}(\exists vx) \Rightarrow (Tr(\exists vx) \equiv \exists t Tr(x(t/v)))]$,

where by $Sent_{\mathcal{L}}(x)$ we mean that x is the Gödel number af an arithmetical (without any occurence of the truth predicate) sentence of the arithmetical language $\mathcal{L}$. Let us note that there are no instances of the induction axiom scheme for formulae of the extended language $\mathcal{L}_{Tr}$ other then those for formulae of the original language of PA among the axioms of CT^{-}. A sentence $(\varphi(0) \wedge \forall x(\varphi(x) \Rightarrow \varphi(s(x)))) \Rightarrow \forall x \varphi(x)$ is an axiom of CT^{-} only if $\varphi \in Frm_{\mathcal{L}}$, i.e. the truth predicate Tr does not occur in φ.[1]

2 Nonstandard Models - Preliminaries

In this section we provvide basic information concerning nonstandard models of arithmetic. .

Let us begin with recalling some basic definitions for this section.

Definition 2 *(i) A* ***signature*** *is a tuple*

$$\sigma = (\{P_i\}_{i \in I}, \{f_j\}_{j \in J}, \{c_k\}_{k \in K})$$

of (non-logical) respectively: predicate, function and constant symbols of the language $\mathcal{L}$, for some index sets I, J and K.
(ii) A ***theory*** *is a set of sentences closed under logical consequence.*
(iii) An ***$\mathcal{L}$-structure (model)***[2] *$\mathcal{M}$ is a tuple $(M, \{P_i^{\mathcal{M}}\}_{i \in I}, \{f_j^{\mathcal{M}}\}_{j \in J}, \{c_k^{\mathcal{M}}\}_{k \in K})$ consisting of the domain (universe), relations, functions and constant elements of $\mathcal{M}$.*
(iv) An $\mathcal{L}$-formula, an $\mathcal{L}$-sentence and an $\mathcal{L}$-theory are, respectively a formula, a sentence and a theory that are defined for the (first-order) language $\mathcal{L}$ (over some signature σ).
(v) A theory Δ is ***satisfiable*** *if and only if there is a model $\mathcal{M}$ such that $\mathcal{M} \models \varphi$ for any $\varphi \in \Delta$.*

Definition 3 *A (full)* ***$\mathcal{L}$-theory of the model*** *$\mathcal{M}$ is a following set of first-order sentences:*
$Th(\mathcal{M}) = \{\varphi \in Sent_L : \quad \mathcal{M} \models \varphi\}$

Definition 4 *Two models ($\mathcal{L}$-structures) $\mathcal{M}_1$ and $\mathcal{M}_2$ are said to be* ***elementarily equivalent****, denoted: $\mathcal{M}_1 \equiv \mathcal{M}_2$, if and only if $Th(\mathcal{M}_1) = Th(\mathcal{M}_2)$.*[3]

Observation 1 *For any models $\mathcal{A}$ and $\mathcal{B}$: $\mathcal{A} \equiv \mathcal{B}$ iff $\mathcal{B} \models Th(\mathcal{A})$.*

[1] The system CT^{-} is called **stratified** since in the axioms $(CT2)$ – $(CT6)$ all the sentences being in the scope of the quantifiers are the sentences of the language $\mathcal{L}$. The axiom $(CT1)$ speaks only of $\mathcal{L}$-sentences because equational theorems are all sentences of $\mathcal{L}$.. A system that is obtained from CT^{-} by adjoining to it all the instances of induction formulated in the ful language $\mathcal{L}_{Tr}$ is called CT.

[2] Sometimes the use of the name *model* is restricted to structures satisfying *given theories.*

[3] Or, equivalently $\mathcal{M}_1 \models \varphi$ iff $\mathcal{M}_2 \models \varphi$ for all $\mathcal{L}$-sentences φ.

Definition 5 *A model $\mathcal{A}$ is a* ***submodel*** *(substructure) of a model $\mathcal{B}$ (denoted $\mathcal{A} \subseteq \mathcal{B}$) if and only if there is an injective function (**embedding**) $g : A \to B$ such that for each $i \in I$, for each $a_1, ..., a_{ar(R_i)} \in A$:*

$$a_1, ..., a_{ar(R_i)} \in R_i^{\mathcal{A}} \quad \Leftrightarrow \quad g(a_1), ..., g(a_{ar(R_i)}) \in R_i^{\mathcal{B}},$$

for each $j \in J$, for each $a_1, ..., a_{ar(f_j)} \in A$:

$$g(f_j^{\mathcal{A}}(a_1, ..., a_{ar(f_j)})) = f_j^{\mathcal{B}}(g(a_1), ..., g(a_{ar(f_j)})),$$

and for each $k \in K$: $g(a_k^{\mathcal{A}}) = a_k^{\mathcal{B}}$.

Definition 6 *Two models $\mathcal{A}$ and $\mathcal{B}$ are said to be* ***isomorphic****, denoted $\mathcal{A} \cong \mathcal{B}$ if and only if there is an embedding $g : A \to B$ such that it is bijection.*

Theorem 1 *Skolem-Löwenheim-Tarski theorem] If a set of first-order formulae Δ over signature σ has an infinite model, then for any cardinal $\mathfrak{m} \geq max\{\aleph_0, |\sigma|\}$, where $|\sigma|$ is the cardinality of the signature σ, there is a model $\mathcal{M} \models \Delta$ such that $|\mathcal{M}| = \mathfrak{m}$.*[4]

Definition 7 *A standard model of arithmetic is a structure*

$$\mathbb{N} = (\omega, +^{\mathbb{N}}, \times^{\mathbb{N}}, s^{\mathbb{N}}, 0^{\mathbb{N}}, <^{\mathbb{N}})$$

such that ω is a set of natural numbers $\{0, 1, 2, ...\}$ and $+^{\mathbb{N}}$, $\times^{\mathbb{N}}$, $s^{\mathbb{N}}$, $0^{\mathbb{N}}$, $<^{\mathbb{N}}$ are respectively addition, multiplication, successor, zero and order on natural numbers.

Definition 8 *An structure $\mathcal{M}$ is called a nonstandard model of arithmetic if $\mathcal{M} \models PA$, but $\mathcal{M} \ncong \mathbb{N}$.*

Theorem 2 (Existence of a nonstandard model of arithmetic) *Let $\mathcal{L}$ over the signature $\sigma = (+, \times, s, 0, <)$ be the language of arithmetic. There is an $\mathcal{L}$-structure $\mathcal{M}$ such that $\mathcal{M} \models PA$ and $\mathcal{M} \ncong \mathbb{N}$.*

[4] Historically speaking, Skolem-Löwenheim-Tarski theorem is a successor of two weaker and older theorems. For full information we state them:

(Downward Skolem-Löwenheim theorem) Any satisfiable theory over signature σ has a model of cardinality less or equal to the cardinality of the set of first-order formulae over σ.

(Upward Skolem-Löwenheim theorem) If a theory over signature σ has an infinite model, then for any cardinal $\mathfrak{m}$ it has a model of cardinality greater or equal to $\mathfrak{m}$.

The oldest protoplast of this group of theorems was simply:

(Skolem-Löwenheim theorem) Any infinite structure $\mathcal{A}$ over at most countable signature contains an at most countable substructure, elementarily equivalent with $\mathcal{A}$.

Proof. Let us adjoin a new constant c to the arithmetical language $\mathcal{L}$ and let $\mathcal{L}^* = \mathcal{L} \cup \{c\}$. Let $\varDelta$ be the following set of $\mathcal{L}^*$-sentences:

$$\varDelta = \{c > 0, c > \overline{1}, c > \overline{2}, ...\} = \{c > \overline{n} : \ n \in \omega\}$$

Now let $\varSigma = PA \cup \varDelta$. We claim that $\varSigma$ is finitely satisfiable. Indeed, there are only finitely many sentences of the form $c > \overline{n}$ in every finite subset $\varSigma_0$ of $\varSigma$. Therefore for any $\varSigma_0$ we take $b = max\{n : c > \overline{n} \in \varSigma_0\}$. It follows that every set $\varSigma_0$ has a model $\mathcal{N} = (\mathbb{N}, a)$ such that $a \in \omega$ and $a = b + 1 = c^{\mathcal{N}}$ (obviously, by the definition: $a > n$, where n is a maximal number such that $c > \overline{n} \in \varSigma_0$). By the Compactness theorem, the theory $\varSigma$ is consistent and satisfiable - it has a model $\mathcal{M}^* = (U, +^{\mathcal{M}^*}, \times^{\mathcal{M}^*}, s^{\mathcal{M}^*}, 0^{\mathcal{M}^*}, <^{\mathcal{M}^*}, c^{\mathcal{M}^*})$. Now let $\mathcal{M} = (U, +^{\mathcal{M}^*}, \times^{\mathcal{M}^*}, s^{\mathcal{M}^*}, 0^{\mathcal{M}^*}, <^{\mathcal{M}^*})$ be the reduct of $\mathcal{M}^*$ to the original arithmetical language $\mathcal{L}$. It holds that $\mathcal{M}^* \models \varSigma$ and therefore $\mathcal{M}^* \models PA$. So, it is clear that

$$\mathcal{M} \models PA.$$ [5]

It also follows that $\mathcal{M} \ncong \mathbb{N}$. Indeed, suppose for the sake of contradiction, that $f : \omega \to U$ is an isomorphism between $\mathbb{N}$ and $\mathcal{M}$. It holds that for every $n \in \omega$:

$$f(n) = \overline{n}^{\mathcal{M}},$$

where $\overline{n}^{\mathcal{M}}$ is a value of a closed term (numeral) $\overline{n}$ in $\mathcal{M}$. But a formula φ of the form $\forall x\, (x > \overline{n} \Rightarrow x \neq \overline{n})$ is a theorem of PA, therefore $\mathcal{M} \models \varphi$ and since by the definition of $\varDelta$ we have that $\forall n \in \omega\ \mathcal{M} \models c > \overline{n}$, it holds that $c^{\mathcal{M}^*} \notin rg(f)$, although $c^{\mathcal{M}^*} \in U$. Therefore f is not surjective and as such cannot be an isomorphism.

Corollary 1 *For any cardinal $\mathfrak{m} \geq \aleph_0$ there is a nonstandard model of arithmetic $\mathcal{M}$ such that $|\mathcal{M}| = \mathfrak{m}$. Hence there is a countable nonstandard model of PA.*

Proof. Immediate by existence of nonstandard models theorem and by the fact that the theory $\varDelta$ (defined in the proof of the existence of nonstandard models) has to have a countable model by Skolem-Löwenheim-Tarski theorem.[6]

We now turn to the question about the order-type of nonstandard models of arithmetic, or as to put it: what do nonstandard models of arithmetic look like?. The results we present here are folklore - an interested Reader may find the omitted proofs e.g. in [1]. For a detailed monograph of on models of PA, see [5] or [7]. For the theorem about the (computational) structure of nonstandard addition and multiplication, additionally to the positions mentioned above see [17].

Definition 9 *The gap of $a \in M$, denoted by $gap(a)$, is the $\mathcal{F}$-gap of a, where $\mathcal{F}$ is the family of* ***all*** *such definable functions, i.e.*

[5] Actually, we even have: $\mathbb{N} \equiv \mathcal{M}$.

[6] By the same argument we obtain that there is a countable nonstandard model of $Th(\mathbb{N})$.

$$\mathcal{F} = \{f : M \to M : f \text{ is definable and } \forall x, y \in M\; x < y \Rightarrow x \leq f(x) \leq f(y)\}.$$

Definition 10 *Recall that an* ***initial segment*** *of $\mathcal{M}$, denoted $\mathcal{I} \subseteq_{end} \mathcal{M}$, is such a subset $\mathcal{I} \subseteq \mathcal{M}$ that is closed* downwards*, i.e. $\forall n \in \mathcal{I}$ and $\forall a \in \mathcal{M}$ if $\mathcal{M} \models a < n$, then $a \in \mathcal{I}$.*

An initial segment $\mathcal{I} \subseteq_{end} \mathcal{M}$ is a ***cut*** *of $\mathcal{M}$, denoted $\mathcal{I} \subseteq_{cut} \mathcal{M}$ if it is nonempty and closed under successor, i.e. $\forall n \in \mathcal{M}\; (n \in \mathcal{I} \Rightarrow s(n) \in \mathcal{I})$.*

A cut $\mathcal{I} \subseteq_{cut} \mathcal{M}$ is a ***proper cut*** *if $\mathcal{I} \neq \mathcal{M}$ (fact: any countable and nonstandard $\mathcal{M} \models PA$ has $2^{\aleph_0}$ proper cuts that are closed under $+, \times$).*

If $\mathcal{M}$ and $\mathcal{N}$ are models of PA, $\mathcal{M}$ is ***cofinal*** *in $\mathcal{N}$, denoted $\mathcal{M} \subseteq_{cf} \mathcal{N}$, if for all $a \in N$ there exists $b \in M$ such that $\mathcal{N} \models a \leq b$.*

Fact 1 *Let $\mathcal{M}$ be an arbitrary model of arithmetic. Then, the function $h : \omega \to |\mathcal{M}|$ defined by the equality:*

$$h(n) = \overline{n}^{\mathcal{M}},$$

is the unique embedding of the standard model $\mathbb{N}$ into $\mathcal{M}$.

Now, we turn to the theorem, that directly characterizes the ordering-type of any nonstandard model of arithmetic. Before that, let us recall that we say that a relation $<$ is ***dense*** if and only if

$$\forall x \forall y\, (x < y \Rightarrow \exists z\, (x < z \wedge z < y)).$$

Lemma 1 (Weak Overspill Principle)) *Suppose that $\mathcal{M}$ is a nonstandard model of arithmetic such that $\forall n \in \omega \;\; \mathcal{M} \models \varphi(n)$ for some $\varphi(x) \in Frm_{\mathcal{L}}$. Then there is a nonstandard element a which also satisfies $\varphi(x)$, i.e. such that $\mathcal{M} \models \varphi(a)$.*

Proof. For the sake of contradiction, suppose that $\forall n \in \omega \;\; \mathcal{M} \models \varphi(n)$ and $\forall a \notin \omega \;\; \mathcal{M} \not\models \varphi(a)$. Then, trivially $\mathcal{M} \models \varphi(0)$ and $\mathcal{M} \models \forall x\, (\varphi(x) \Rightarrow \varphi(s(x)))$. But $\mathcal{M}$ has to satisfy the instance of induction axiom for φ, so $\mathcal{M} \models \forall x \varphi(x)$. Contradiction.

Definition 11 *For any model $\mathcal{M}$, an* ***initial segment*** *of $\mathcal{M}$, denoted $\mathcal{I} \subseteq_{end} \mathcal{M}$, is such a subset $\mathcal{I} \subseteq \mathcal{M}$ that is closed* downwards*, i.e. $\forall n \in \mathcal{I}$ and $\forall a \in \mathcal{M}$ if $\mathcal{M} \models a < n$, then $a \in \mathcal{I}$.*

Theorem 3 *For any nonstandard model $\mathcal{M} = (|\mathcal{M}|, <^{\mathcal{M}})$, the ordering $<^{\mathcal{M}}$ is isomorphic with the ordering $(\omega + \mathbb{Z} \cdot \eta, <)$, where η is an order-type of some dense, linear order (DLO).*

Corollary 2 *For any nonstandard, countable model $\mathcal{M} = (|\mathcal{M}|, <^{\mathcal{M}})$, the ordering $<^{\mathcal{M}}$ is isomorphic with the ordering $(\omega + \mathbb{Z} \cdot \mathbb{Q}, <)$.*

Proof. Immediate by Cantor's theorem that any countable structure satisfying the properties of dense, linear order is isomorphic to $(\mathbb{Q}, <)$.

Let us now ask a question: how many non-isomorphic countable nonstandard models of arithmetic are there? The following folklore theorem answers this question:

Theorem 4 *There are exactly $2^{\aleph_0}$ non-isomorphic countable models of PA.*

Proof. It is obvious that there is at most $2^{\aleph_0}$ non-isomorphic countable models of PA. We will show then that there is also at least $2^{\aleph_0}$ such models and by Cantor-Bernstein-Schröder theorem[7] we will conclude that there are exactly $2^{\aleph_0}$ of them. Suppose $p_0 = 2$, $p_1 = 3$, $p_2 = 5$, ... are the standard primes and let $S \subseteq \omega$ be arbitrary. We adjoin a new constant c to our language $\mathcal{L}$ and consider a following theory:

$$\Delta = PA \cup \{c > \overline{n} : n \in \omega\} \cup \{\forall x \neg(\overline{p_k} \cdot x = c) : k \notin S\} \cup \{\exists x (\overline{p_k} \cdot x = c) : k \in S\}.$$

Every finite $\Delta_0 \subseteq \Delta$ is contained in one of the theories Δ_m defined as follows:

$$\begin{gathered} \Delta_m = PA \ \cup \ \{c > \overline{n} : \ n < m\} \ \cup \\ \cup \ \{\forall x \neg(\overline{p_k} \cdot x = c) : \ k \notin S \wedge k < m\} \ \cup \\ \cup \ \{\exists x \ (\overline{p_k} \cdot x = c) : \ k \in S \wedge k < m\} \end{gathered}$$

for some fixed $m \in \omega$. Let q be any prime in ω such that $q > m$ and let r be defined for $k \in S$ as follows:

$$r = q \cdot \prod_{k<m} p_k.$$

Obviously for any m we have a model $\mathcal{N}_m = (\mathbb{N}, r) \models \Delta_m$, where $r = c^{\mathcal{N}_m}$. Thus Δ is finitely satisfiable and hence by Compactness has a model. Let $\mathcal{M}_c \models \Delta$ and let $\mathcal{M}$ be the reduct of $\mathcal{M}_c$ to the original language $\mathcal{L}$ and let $a \in |\mathcal{M}|$ be a nonstandard element such that $a = c^{\mathcal{M}_c}$. Then we have:

$$S := S_a = \{n \in \omega : \ \mathcal{M} \models \ \exists x \ (\overline{p_n} \cdot x = a)\}$$

and we say that S_a is coded by a. We thus have shown that **any** set $S \subseteq \omega$ is coded by some nonstandard a in some nonstandard, countable model $\mathcal{M} \models PA$. We will now show that there are at least $2^{\aleph_0}$ non-isomorphic countable models $\mathcal{M}_i$ of PA. Suppose for the sake of contradiction that there are $\mathfrak{m} < 2^{\aleph_0}$ such models and let i range over an index set I of cardinality $\mathfrak{m}$. Then the number of subsets $S \subseteq \omega$ coded by some a in some $\mathcal{M}_i$ is:

$$\leq \sum_{i \in I} card(\mathcal{M}_i) = \mathfrak{m} \times \aleph_0 = max\{\mathfrak{m}, \aleph_0\},$$

which is strictly less than $2^{\aleph_0}$. But since there are $2^{\aleph_0}$ subsets of ω it is a contradiction which ends the proof.[8]

[7] Cantor-Bernstein-Schröder theorem states that for any sets A and B, if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

[8] By the same argument we may obtain that there are exactly $2^{\aleph_0}$ non-isomorphic countable models of $Th(\mathbb{N})$.

3 Satisfaction Classes

Definition 12 *Let $\mathcal{M}$ be a model of PA. A set $S \subseteq |\mathcal{M}|$ is a* ***full satisfaction class*** *for $\mathcal{M}$ if and only if $(\mathcal{M}, S) \models CT^-$.*

Although we will not prove this fact here, it is worth noting that not each countable model of PA admits a full satisfaction class.

Definition 13 *A set p of the formulae of the language $\mathcal{L}_{\mathcal{M}}$ (i.e. the language $\mathcal{L}$ extended with a constant name for every element of the model $\mathcal{M}$) with exactly one free variable x is* ***finitely satisfied*** *in $\mathcal{M}$ if and only if for any finite $q \subset p$ there exists an $a \in |\mathcal{M}|$ such that for any $\varphi(x) \in q$ $\mathcal{M} \models \varphi(a)$. A* ***type*** *over a model $\mathcal{M}$ is a finitely satisfied set of formulae of the form $\varphi(x, b)$ with exactly one free variable x and at most one parameter $b \in M$. A type p over $\mathcal{M}$ is* ***recursive*** *if and only if the set of codes of formulae $\varphi(x, y)$ such that $\varphi(x, b) \in p$ is recursive. A type p over $\mathcal{M}$ is (globally)* ***realised*** *if and only if there exists an $a \in M$ such that for any $\varphi(x, b) \in p$ $\mathcal{M} \models \varphi(a, b)$. Model $\mathcal{M}$ of PA is* ***recursively saturated*** *if and only if each recursive type over $\mathcal{M}$ is realised.*

Theorem 5 (**Lachlan's theorem**, A. Lachlan [12], also see R. Kaye [5], p. 150, pp. 228-233) *If a nonstandard model $\mathcal{M} \models (PA)$ admits a full satisfaction class, then $\mathcal{M}$ is recursively saturated.*

Theorem 6 ((KKL)H. Kotlarski and S. Krajewski and A. Lachlan [9], J. Barwise and J. Schlipf [2], A. Enayat, and A. Visser [3])
If a countable model $\mathcal{M} \models (PA)$ is recursively saturated, than it admits a full satisfaction class.

Therefore, for a countable model of arithmetic, it is equivalent to admit a full satisfaction class (i.e. satisfy the formal theory of compositional truth) and to be recursively saturated. Therefore the project can be thought of as an investigation into structure of possible interpretations for theory of compositional, arithmetical truth. It needs to be underlined that the purpose of our research is to examine model-theoretic but purely arithmetical properties of models admitting satisfaction classes. In particular, we study various substructures of recursively saturated models of PA.

4 Cofinal Extensions, Gaps and Cuts

What we aim to study here are certain logical properties of cofinal extensions of submodels (actually, cuts) of models of PA. It might be said in doing so we follow the advice of Smorynski expressed in [14]:

> A relatively neglected aspect of the study of nonstandard models of arithmetic is the study of their cofinal extensions. These extensions

> certainly do not present themselves to the intuition as readily as do their more popular cousins the end extensions; but they are not exactly shrouded in mystery or unnatural objects of study either. They are equal partners with end extensions in the construction of general extensions of models; they offer both special advantages and disadvantages worthy of our interest; and, occasionally, they are useful in understanding the generally more simply behaved end extensions. Cofinal extensions deserve more attention than they have traditionally received.

First-order theories of pairs (N, M), where $N \models PA$ and M is an elementary cofinal submodel of M reveal great diversity and demand systematic study. The case of models admitting satisfaction classes is of particular interest in this respect: all countable recursively saturated models of PA have continuum many nonisomorphic cofinal submodels, and after acknowledging the variety of the abovementioned pairs for N being countable recursively saturated, the next goal is to consider isomorphism types and first-order theories for pairs of models (N, M) for a fixed countable recursively saturated model N and a fixed isomorphism type of M. The method that has already been shown quite effective in this direction is the method of *gaps* (also called *skies*). We present briefly the gap terminology and explain why it is useful.v

Skolem terms, also called simply definable functions[9], are paramter-free definable and PA-provably total functions. Let $\mathcal{M}$ be a nonstandard model of arithmetic and let $\mathcal{F}$ be some family of Skolem terms $f : M \to M$ such that $\forall x, y \in M\, x < y \Rightarrow x \leq f(x) \leq f(y)$. There is a partition of M into sets, which we call $\mathcal{F}$-gaps. For any $a \in M$, $gap_{\mathcal{F}}(a)$ is the smallest set $C \subseteq M$ such that $a \in C$ and:

$$\forall b \in C\, \forall f \in \mathcal{F}\, \forall x \in M\;\; b \leq x \leq f(b) \;\vee\; x \leq b \leq f(x)) \Rightarrow x \in C.$$

This is a natural generalization of an idea of partitioning the universe of a nonstandard model into $\mathbb{Z}$-blocks around each element (then, each such block is $gap_{\mathcal{F}}(a)$ for some a, where $\mathcal{F}$ consists only of the successor function s).

Every model $\mathcal{M}$ has the least gap, the $gap(0)$. Let $A \subseteq M$. Then, we denote $sup(A) = \{x \in M : \exists y \in A\, x \leq y\}$. If for some $a \in M$, $M = sup(gap(a))$, then we call $gap(a)$ the **last gap of** $\mathcal{M}$. A model with a last gap is called **short**. If $\mathcal{M} \preceq_{cut} \mathcal{N}$ (i.e. $\mathcal{M}$ is an elementary cut of $\mathcal{N}$), we say that $\mathcal{M}$ is **short elementary cut** of $\mathcal{N}$ if $\mathcal{M}$ is short - in other words, if by $Scl(a)$ we denote the set of all $t(a)$ such that t is a Skolem term of PA, $\mathcal{M}$ is short if there is such an element $a \in M$ that its Skolem closure in $\mathcal{M}$ is cofinal in $\mathcal{M}$, i.e. for all $x \in M$ there is $b \in Scl(a)$ such that $x <_{\mathcal{M}} b$. An elementary cut is **coshort** if $\mathcal{N} \setminus \mathcal{M}$ has the least gap, i.e. there is $a \in N \setminus M$ s.t. $M = inf(gap(a))$, where $inf(A) = \{x \in M : \forall y \in A\, x \leq y\}$.

Now, to clarify the gap terminology, if we put:

- $\mathcal{M}(a) = sup(Scl(a))$, and

[9] with a slight abuse of terminology that is unimportant to our investigations

- $\mathcal{M}[a] = \{b \in M : \forall t \in Scl(b)\, t(b) < a\}$,

then the set $[a) = \mathcal{M}(a) \setminus \mathcal{M}[a]$ is exactly the $gap(a)$. It can be shown that $\mathcal{M}(a)$ is the smallest elementary cut of $\mathcal{M}$ containing a, and that $\mathcal{M}[a]$ is empty if and only if every elementary cut of $\mathcal{M}$ contains a. Gap terminology is particularly useful in the study of recursively saturated models of PA (see e.g. [7] for a reference to many methods and properties).

Let $\mathcal{N}$ be a countable recursively saturated model of PA and let $\mathcal{E}(\mathcal{N})$ denote the family of its elementary submodels. If every complete type realised in a countable model $\mathcal{M} \models PA$ is also realised in $\mathcal{N}$, then $\mathcal{M}$ can be elementarily embedded into $\mathcal{N}$. This implies that there are uncountably many nonisomorphic submodels in any $\mathcal{N} \models PA$. Choice for elementary cuts is more limited: if $\mathcal{M} \in \mathcal{E}(\mathcal{N})$ is a cut, then either

- $\mathcal{M}$ is tall, and in this case $\mathcal{M}$ is isomorphic to $\mathcal{N}$,
  or
- $\mathcal{M}$ is short, and then $\mathcal{M}$ is one of the countably many nonisomorphic elementary cuts of the form $sup(Scl(a))$.

Thus, it is fair to state that short cuts in models of arithmetic are cofinal extensions of canonical subsets of the model, i.e. they are supremas (or downsets, which is equivalent in the case of the structures we consider, i.e. the models of arithmetic) of a Skolem closure of a given element.

So, the question rises: how many short elementary cuts are there?

Theorem 7 (Smorynski, [15]) *Every countable recursively saturated model of PA has infinitely many pairwise nonisomorphic short elementary cuts.*

5 The Isomorphism Problem for Pairs

One of the interesting and natural questions concerning *pairs* for countable recursively saturated models of arithmetic and its cuts is the following *big* question of our particular interest:

Let $\mathcal{M} \models PA$ be a countable recursively saturated model and let $a, b \in M$. Suppose that $(\mathcal{M}, \mathcal{M}(a)) \equiv (\mathcal{M}, \mathcal{M}(b))$. Does it follow that $(\mathcal{M}, \mathcal{M}(a)) \cong (\mathcal{M}, \mathcal{M}(b))$?

Another way to put it is: under what conditions, does the identity of theories of such pairs imply their isomoprhism? It is known that the answer to the *big* question above is negative, if $\mathcal{M}(a)$ and $\mathcal{M}(b)$ in question are coshort, as shown by R. Kossak and J. Schmerl in [8]. However, it remains open (and is considered to be difficult) what is the answer for the case in which $\mathcal{M}(a)$ and $\mathcal{M}(b)$ are short elementary cuts of $\mathcal{M}$.

The exact state of the art for the coshort case is as follows:

Theorem 8 (R. Kossak, H. Kotlarski, [6]) *Let $\mathcal{M}$ be a countable recursively saturated model of PA. Then for all $a, b \in \mathcal{M} \setminus \mathcal{M}(0)$, the strutures $(\mathcal{M}, \mathcal{M}[a])$ and $(\mathcal{M}, \mathcal{M}[b])$ are elementarily equivalent*

Definition 14 (Set of complete types realised by a set) *For any $A \subseteq M$, $Tp(A) = \{tp^{\mathcal{M}}(a) : a \in A\}$*

Fact 2 *Let $\mathcal{M} \models PA$ be countable recursively saturated. Then, for all $a, b \in M$, either $Tp(gap(a)) = Tp(gap(b))$ or $Tp(gap(a)) \cap Tp(gap(b)) = \emptyset$.*

Theorem 9 (R. Kossak, J. Schmerl, [8]) *1. Let $\mathcal{M} \models PA$ be a countable recursively saturated model. There are infinite sets L and U of gaps such that for distinct gaps γ and γ' in either L or U, $Tp(\gamma) \cap Tp(\gamma') = \emptyset$.*

2. If γ and γ' are the least gaps in $\mathcal{M} \setminus \mathcal{K}$ and $\mathcal{M} \setminus \mathcal{K}'$, respectively, such that $Tp(\gamma) \cap Tp(\gamma') = \emptyset$, then $(\mathcal{M}, \mathcal{K}) \ncong (\mathcal{M}, \mathcal{K}')$.

3. There are infinitely many pairs of c.r.s. models $\mathcal{M} \models PA$ and coshort elementary cuts $\mathcal{K}$ and $\mathcal{K}'$ with (distinct) least gaps γ and γ' , respectively, such that $Tp(\gamma) \neq Tp(\gamma')$.

Thus: there are infinitely many elementarily equivalent and pairwise nonisomorphic pairs $(\mathcal{M}, \mathcal{K})$ with $\mathcal{K}$ being coshort.

To provide a partial answer for the short case, we will use the following result of Smorynski:

Theorem 10 (Smorynski [15]) *Let $\mathcal{M} \models PA$ be a countable recursively saturated model and let $\mathcal{K} \in \mathcal{E}(\mathcal{M})$ be short (i.e. having a last gap). Then the following are definable without parameters in $(\mathcal{M}, \mathcal{K})$:*

1. *$\mathbb{N}$,*
2. *the truth definition for $\mathcal{K}$,*
3. *the last (max) gap.*

Since it is not hard to prove the equivalence that there exists an automorphism of such $\mathcal{M}$ if and only if $tp(a) = tp(b)$ (in the purely arithmetical language $\mathcal{L}$), where $tp(a) = \{\varphi(x) : \mathcal{M} \models \varphi(a)\}$ is the set of formulae satisfied in $\mathcal{M}$ by $a \in M$, the natural way to proceed is to consider the definable sets in $(\mathcal{M}, \mathcal{M}(a))$ and complete types realized in the last gap of $\mathcal{M}$. We might then first ask under what circumstances there is an element $c \in gap(a)$ such that $tp(c) \in Def(\mathcal{M}, \mathcal{M}(a))$ for $\mathcal{M}$ being a countable recursively saturated model of PA. We use results of Smorynski from [14] and work with gaps and standard systems $SSy(\mathcal{M})$ of $\mathcal{M}$, i.e. the family of all subsets of $\mathbb{N}$ that are coded in $\mathcal{M}$[10].

Definition 15 (Standard System) *1. Let $\mathcal{M} \models PA$ and let $a \in M$. **The set coded by** a **in** $\mathcal{M}$ is the set*

$$S_a = \{n \in \omega : \ \mathcal{M} \models \exists x\, (\overline{p_n} \cdot x = a)\}.$$

*2. $\mathcal{M}$ **codes** $S \subseteq \mathbb{N}$ iff there is $a \in M$ such that a codes S in $\mathcal{M}$.*

[10] It turns out that the standard system tells you a lot about the model; for example, any two countable recursively saturated models of the same completion of PA with the same standard system are isomorphic.

3. ***The standard system of*** *$\mathcal{M}$, denoted $SSy(\mathcal{M})$, is the set of all subsets of $\mathbb{N}$ that are coded in $\mathcal{M}$, i.e.*

$$SSy(\mathcal{M}) = \{S \subseteq \mathbb{N} : \exists a \in M\ S = \{n \in \omega :\ \mathcal{M} \models \exists x\, (\overline{p_n} \cdot x = a)\}.$$

Definition 16 *Let $\mathcal{M} \models PA$. Then $Def(\mathcal{M})$ denotes the family of sets that are* ***definable (with parameters) in*** *$\mathcal{M}$.*

When the standard system of a given model is relatively simple, than isomorphism of pairs follows.

Theorem 11 (Tin Lok Wong, MTG) *Let $\mathcal{M} \models PA$ be a countable recursively saturated model and let $a, b \in M$. Suppose that $(\mathcal{M}, \mathcal{M}(a)) \equiv (\mathcal{M}, \mathcal{M}(b))$ and assume $\mathcal{M}(a)$ and $\mathcal{M}(b)$ are short. If $SSy(\mathcal{M}) \subseteq Def(\mathbb{N})$, then $(\mathcal{M}, \mathcal{M}(a)) \cong (\mathcal{M}, \mathcal{M}(b))$.*

Proof. The proof uses the relativisation of $SSy(\mathcal{M})$ to definability in the standard model $\mathbb{N}$ - it enables us to describe the appropriate type.

By the fact that $\mathcal{M}$ is recursively saturated, the computable set of formulae that $tp(a)$ consists of is satisfied. Observe that the set of Gödel numbers of the formulae in $tp(a)$ is therefore a coded subset of $\mathbb{N}$, thus we have $tp(a) \in SSy(\mathcal{M})$.

Hence, $tp(a)$ is definable, i.e. $tp(a) \in Def(\mathbb{N})$ which follows from our main assumption.

Suppose $\varphi(x)$ defines $tp(a)$ in $\mathbb{N}$. Then:

$$(\mathcal{M}, \mathcal{M}(a)) \models \exists v \in maxgap\, \forall x(\varphi(x) \Rightarrow Sat_{\mathcal{M}(a)}(x, v)),$$

which follows from Smorynski's result, since the last gap is definable, as well as the satisfaction predicate for the short elementary cut in question. Therefore, since the models in question are elementarily equivalent, i.e. $(\mathcal{M}, \mathcal{M}(a)) \equiv (\mathcal{M}, \mathcal{M}(b))$, we have that the same satisfaction predicate works for $\mathcal{M}(b)$ and that $tp(a) = tp(b)$ which guarantees the existence of an isomorphism: $(\mathcal{M}, \mathcal{M}(a)) \cong (\mathcal{M}, \mathcal{M}(b))$.

The conceptual import of the result is that taking a nonstandard model of compositional truth such that all its coded sets are already definable in the standard model, we are able to identify isomorphic short elementary cuts that are canonical cofinal extensions of a subset of the model just by looking at the arithmetical theory of both pairs considered.

References

1. Z. Adamowicz and P. Zbierski, *Logic of mathematics*, John Wiley & Sons, New York, 1997.
2. J. Barwise and J. Schlipf, *An introduction to recursively saturated and resplendent models*, **Journal of Symbolic Logic** 41 (1976), 531-536.

3. A. Enayat and A. Visser, *New constructions of satisfaction classes*, in *Unifying the Philosophy of Truth*, ed. T. Achourioti, H. Galinon, J. Martinez Fernandez, K. Fujimoto, 321-335.
4. F. Engström, *Satisfaction classes in nonstandard models of first-order arithmetic*, Chalmers University of Technology and Göteborg University, Göteborg, 2002.
5. R. Kaye, *Models of Peano Arithmetic*, Oxford University Press, Oxford, 1991.
6. R. Kossak and H. Kotlarski, *On extending automorphisms of models of Peano Arithmetic*, **Fundamenta Mathematicae** 149/3 (1996), 245 - 263.
7. R. Kossak and J. Schmerl, *The structure of models of Peano Arithmetic*, Clarendon Press, Oxford, 2006.
8. R. Kossak and J. Schmerl, *On Cofinal Submodels and Elementary Interstices*, **Notre Dame Journal of formal Logic** 53 (2012), 267 - 287.
9. H. Kotlarski and S. Krajewski and A. Lachlan, *Construction of satisfaction classes for nonsatndard models*, **Canadian Mathematical Bulletin** 24 (1981), 283-293.
10. H. Kotlarski, *Full satisfaction classes: a survey*, **Notre Dame Journal of Formal Logic** 32 (1991), 573-579.
11. S. Krajewski, *Nonstadard satisfaction classes*, in *Set Theory and Hierarchy Theory*, ed. W. Marek, M. Srebrny, A. Zarach, Heidelberg, 1976.
12. A. Lachlan *Full satisfaction classes and recursive saturation*, **Canadian Mathematical Bulletin** 24 (1981), 295-297.
13. A. Robinson *On languages based on on-standard arithmetic*, **Nagoya Mathematical Journal** 22 (1963), 83 - 107.
14. C. Smorynski, *Cofinal extensions and nonstandard models of arithmetic*, **Notre Dame Journal of formal Logic** 2 (1981), 133 - 144.
15. C. Smorynski, *Elementary extensions of recursively saturated models of arithmetic*, **Notre Dame Journal of formal Logic** 2 (1981), 193 - 203.
16. C. Smorynski, *Recursively saturated nonstandard models of arithmetic*, textbfJournal of Symbolic Logic 46 (1981), 259 - 286.
17. S. Tennenbaum, *Non-archimedean models for arithmetic*, **Notices of American Mathematical Society** 6 (1959), 270.

Exploring Tractability in Finitely-Valued SAT Solving*

Nika Pona

Vienna University of Technology
nika@logic.at

Abstract. In this paper I describe the progress, preliminary results and future work directions of a project of implementing a many-valued SAT solver based on a generalization of algorithms used in modern Boolean SAT solvers. Mimicking Boolean SAT solvers minimizes the algorithm-design and implementation challenges related to such a task, since many ideas can be easily adapted to the many-valued setting. Experimental results show that even on the early stages of the development a many-valued solver can perform better on some problems than modern Boolean SAT solvers.

1 Introduction and Motivation

The starting idea of the project was to see whether the current many-valued solvers could be improved using the theoretical results in complexity of finitely-valued logics [5]. The research of the state of the art in the field showed that there are no complete many-valued SAT solvers available, and that the most common approach to solve the problems modelled as many-valued formulae is to reduce them to Boolean SAT[1] or Satisfiability Modulo Theory (SAT with Linear Arithmetic Theory) [2][2]. Previously some complete many-valued solvers were implemented and the results seemed to be promising [7] [8], but the projects were discontinued and the software is not available any more. Thus the task became to implement a many-valued SAT solver first.

1.1 Why Many-Valued SAT?

SAT solving has enjoyed a lot of success in the last two decades due to an organized effort of the growing community of researchers. Since any finitely-valued logic formula can be efficiently mapped to an equisatisfiable Boolean logic formula by encoding the information about the many-valued domain with additional constraints (cf.[3]) it may seem that investing time into a separate many-valued solvers is superfluous. There are two reasons to think that such an implementation effort can be interesting:

* This project is supported by the Austrian Science Fund (FWF): I836-N23.

[1] For the description of the most common encodings and their properties see references here: http://bach.istc.kobe-u.ac.jp/sugar/

[2] There is a solver available online that uses this approach: http://www.iiia.csic.es/~amanda/files/2012/NiBLoS.zip.

Many-valued SAT as generalized SAT Investigating many-valued logics proved to be useful in complexity and proof theory, where Boolean logic is seen as a special case with two truth values. One can expect that similar results can be achieved with respect to algorithms for the SAT problem. The conflict-driven DPLL algorithms that are the basis of all modern SAT solvers generalize easily to the many-valued setting: the literal watching scheme, Unique Implication Point learning method and counter-based decision heuristics can be implemented in basically the same way as in a SAT Solver (more on this below). This means that the effort required for designing and implementing a many-valued solver is *relatively* small; at the same time, looking at the Boolean SAT algorithms as special cases of a more general scheme can provide some useful insights into SAT solving.

CSP and many-valued SAT Another reason to look into many-valued SAT is that it can be seen as an intermediate language between Constraint Satisfaction Problems and SAT or even a better alternative to SAT when it comes to CSP solving. We know that CSP can be efficiently translated into SAT, and this fact was used in the CSP community to develop solvers. For instance, the CSP solver Sugar used by Scala constraints language: `http://bach.istc.kobe-u.ac.jp/sugar/`, `http://bach.istc.kobe-u.ac.jp/copris/`, furthermore several solvers of the MiniZinc CSP Challenge 2015 are based on translations to SAT as well: see `http://www.minizinc.org/`. This is the easiest, but not necessarily the most effective approach, since the encodings can become quite big and, most importantly, the structure of the formula is lost and many unnecessary propagations are made. One can translate CSP into a many-valued CNF formula by representing no-goods of every constraint as a clause. Such translation preserves the structure (domains) of the original problem, thus it may be more efficient to use a many-valued SAT Solver as a the back-end of a generic CSP solver. Below I will provide an example that supports this claim.

1.2 Overview

The main point of this presentation is to show that creating a competitive solver for many-valued logic is not as challenging as it may seem, and given the advantages of many-valued modelling it is a potentially fruitful direction of research. To the paper I attach the core solver that implements several versions of a basic conflict-driven algorithm with a resolution-based learning procedure: `https://github.com/akinanop/mvl-solver`[3].

After the basic definitions of Section 2, I first provide empirical results from testing the implemented solver and some theoretical remarks on the advantages of many-valued solving (Section 3), since they provide motivation for the implementation task undertaken. In particular, I give an example where modelling a problem as a many-valued formula and solving it directly with a many-valued

[3] For more details on the actual implementation, see the readme file and the wiki pages of the project.

solver is significantly (one-two orders of magnitude just in terms of solving time) more efficient than formalizing it as a Boolean formula and using a SAT solver, even a competitive one. Then in the Section 4 I describe the general idea of the implementation and finally in Section 5 point to further directions of development of the project.

2 Definitions

Definition 1 (Many-valued SAT). *A many-valued SAT problem $P = (V, D, C)$ is specified by a finite set V of variables, collection of sets (domains) D and the set C of clauses. Each variable $v \in V$ has an associated finite domain $dom(v) \in D$. To solve the many-valued SAT problem P is to determine whether there is an interpretation that satisfies all clauses in C.*

Definition 2 (Literal, clause). *A clause is a finite set of literals. A literal is an expression of the form $v = x$ or $v \neq x$, where $v \in V$ and $x \in dom(v)$. A literal of the form $v = x$ is called positive; a negative literal is of the form $v \neq x$.*

Alternatively, one could consider many-valued literals of the form $v \in A$ with $A \subseteq dom(v)$. The former representation is closer to Boolean SAT, thus permits easier adaptation of the Boolean SAT algorithms. In particular, the input to the many-valued SAT Solver can be given in a format similar to DIMACS in Boolean SAT[4]. Although the second formulation can give some advantages to many-valued SAT, it departs from Boolean SAT and may provide additional implementation challenges, thus I leave its exploration for later. See, for instance [8] for such a formulation.

Definition 3 (Interpretation, model). *An interpretation is a function mapping each variable $v \in V$ to a value from $dom(v)$. An interpretation I satisfies a positive literal $v = x$ if $I(v) = x$, and satisfies a negative literal $v \neq x$ if $I(v) \neq x$. An interpretation satisfies a clause if it satisfies at least one of the literals from the clause.*

3 Modelling Advantage of Many-valued SAT

I use the developed solver to show that solving some problems directly as many-valued problems can have a significant advantage. The authors of the previous attempts to create a complete many-valued solver argued that overall their solvers performed better than the Boolean SAT solvers [7] [8]. However, these projects date back 13 and 6 years respectively, thus it is possible that the progress in SAT solving of the last decade made these results obsolete. Below I show an example of what can be called an intrinsic advantage of the many-valued formulation of a problem: in this case, despite the implementation advances in Boolean SAT, the

[4] For exact specification see here: `https://github.com/akinanop/mvl-solver/wiki/Extended-DIMACS-format`

many-valued solver still performs better. In particular, I compare the developed many-valued solver to `minisat` and some competitive solvers on the Pigeonhole problem and n-queens problem. Moreover, I show that encoding a problem into Boolean SAT via many-valued formulation already gives an advantage in the search.

3.1 Pigeonhole problem

Pigeonhole problem (PHP) is a famous unsatisfiable problem, since despite it's easy formulation: "it is impossible to fit n pigeons into $n-1$ holes, such that each hole contains exactly one pigeon", its unsatisfiability is known to be difficult to prove via automatic means[5]. I consider the following encodings of the PHP:

SAT The Boolean SAT PHP is usually formulated as a CNF formula with variables x_{ij} for each pair $i \in [n]$ and $j \in [n-1]$ and with two types of clauses for all $m \in [n-1]$:

1. $\bigvee_i x_{im}$ for $i \in [n]$;
2. $\neg x_{km} \vee \neg x_{lm}$ for $k \neq l \in [n]$

MVL The many-valued SAT PHP consists of n variables of domain $n-1$. Domain declaration express the condition that each pigeon should be placed in some hole, and the clauses $k \neq j \vee l \neq j$ for $k \neq l \in [n]$ and $j \in [n-1]$ express the condition that no two pigeons should be placed in the same hole.

MVL-SAT Additionally I consider a different Boolean SAT formulation of the PHP – created by automatically translating a many-valued PHP into a Boolean formula using *linear encoding* described in [6]. Replace each (negative) literal of a many-valued PHP with a (negated) boolean variable. As in SAT encoding add clauses of type 1. Furthermore, for each many-valued variable v, introduce $|dom(v)|-1$ new Boolean variables v_i which will be used to enforce the property that at most one value has to be assigned to the variable. This will introduce only linear increase in the size of the original problem, unlike if one does it naively via binary inequalities $v \neq i \vee v \neq j$. For $i \in \{2, \ldots, |dom(v)|-1\}$ add:

1. $\neg v_{i-1} \vee v \neq i$;
2. $v \neq i \vee v_i$;
3. $\neg v_i \vee v_{i-1} \vee v = i$;
4. $\neg v_1 \vee v = 1$.

Below are the characteristics of these encodings:

[5] Resolution-based proofs of unsatisfiability of pigeonhole problem have exponential lower bounds. Pure CDPLL algorithms for SAT are not stronger than resolution, thus this result carries over. However, this can be improved by introducing the so-called symmetry breaking clauses [1]. For instance, one of the winning solvers in 2015 `lingeling` uses symmetry-detecting preprocessing and thus solves PHP instances fast: `http://fmv.jku.at/papers/BiereLeBerreLoncaManthey-SAT14.pdf`.

Table 1. Number of variables and clauses on PHP with $n = 10 \ldots 15$

MVL		SAT		MVL-SAT	
variables	clauses	variables	clauses	variables	clauses
10	405	90	415	170	725
11	550	110	561	209	946
12	726	132	738	252	1206
13	936	156	949	299	1508
14	1183	182	1197	350	1855
15	1470	210	1485	405	2250

Below you can see that `mvl-solver` needs less time then `minisat` on both Boolean formulations of PHP[6]. On $n = 15$ neither `minisat`, nor modern 2014-2015 winner solvers `glucose` and `COMiniSatPS`[7] terminated within 24 hours, whereas `mvl-solver` with both heuristics[8] was finished within 10-17 hours. Since the architecture of the `mvl-solver` is quite basic and not quite efficient yet (in particular, the propagation is very slow – on big satisfiable graph coloring instances where only extensive propagation is needed `mvl-solver` performs slowly compared to `minisat` that finishes instantly), one can make the conclusion the difference lies in the modelling advantage of the many-valued SAT.

Table 2. Times (s) on PHP with $n = 10 \ldots 15$

	`minisat`		`mvl-solver`		`COMiniSatPS`
n	MVL-SAT	SAT	BK	VSIDS	SAT
15	t/o	t/o	17hrs	10hrs	t/o
13	19hrs	t/o	3239	999	13hrs
12	1061	1624	414	170	450
11	49	82	44	26	24
10	3	6	4	3	3

Below I also provide other statistics on this problem: the number of conflicts is significantly smaller for the `mvl-solver`, which is responsible for its better performance, since propagation is slow due to the experimental implementation.

[6] All tests are done on a machine with Intel Core i3-6100 CPU @ 3.70GHz × 4 processor and 7.7 GB memory.

[7] For the results of the 2015 SAT Race see here: `http://baldur.iti.kit.edu/sat-race-2015/index.php?cat=results`

[8] BK chooses the literal that maximizes propagation effect based on currently unsatisfied clauses; VSIDS chooses the literal that occurs in more clauses, then counts for all the literals in the theory are divided by 2 after a learned clause is added to the clause set.

From this table one can see the second interesting result: the decrease in all indicators for MVL-SAT encoding compared to the SAT encoding: the additional constraints added from MVL encoding help trim the search space considerably. This confirms that exploiting structural information through MVL-encoding can be beneficial on difficult, but structured problems[9].

Table 3. Other statistics on PHP with $n = 10$

	MVL-SAT	SAT	MVL (VSIDS)
Restarts	1023	2047	0
Conflicts	472432	1034642	1793
Decisions	522643	1243538	1793
Propagations	7077471	12935371	50778
CPU time (s)	3	6	3

3.2 N-queens

I also compared the performance of `minisat` and `mvl-solver` on the n-queens problems for $n = 4 \ldots 70$, which are typically not very difficult, albeit large, satisfiable problems. See Figure 1. below for the results of the tests[10]. Here the advantages are not as clear as in the case of the pigeonhole problem (overall the solver perform worse time-wise), but despite of this some observations can be made. The number of conflicts in `minisat` is quite small (less than 300 on any instance), however, on more instances `mvl-solver` "got lucky" and had even smaller number of conflicts or no conflicts at all; on the other hand, on some cases it got stuck and needed up to 10-20 times more backtracks. In each case `minisat` performed 2-4 restarts, which suggests that this could also be useful in `mvl-solver` to avoid the bottlenecks. Then the performance could become better overall, since in a many-valued case it is easier to guess a solution to these problems. Some preliminary testing showed that on cases were `mvl-solver` got stuck, restarts do improve the situation, however, more work is needed to provide a stable improvement on all instances using restarts.

4 Algorithms and Implementation

Currently there are no complete many-valued solvers available to the public that are not based on translations to SAT or SMT. Thus the main task of the

[9] The creators of `glucose` complain that most solvers were created with the aim of improving propagation (in order to learn more clauses faster), but this is not so important for difficult cases. Thus they look into the structure of the problems and learned clauses, hence the idea of useful strong "glue" clauses [4].

[10] 8 instances require more than 800 backtracks.

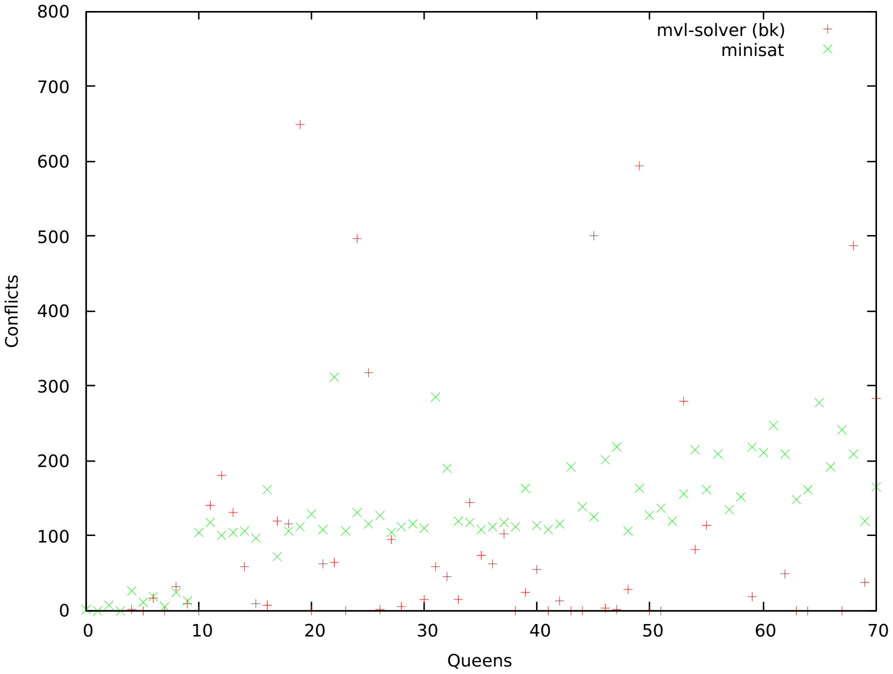


Fig. 1. `minisat` and `mvl-solver` on n-queens

project was to develop such a solver. I reused parts of the open source software (written in C++) created by a Master Student at the University of Minnesota in 2005: `http://www.d.umn.edu/~lalx0004/research/`. It contained some severe algorithmic and implementation mistakes, but provided a good starting point. Thus I kept the input/output part of the solver as well as most of the data structures, but I implemented a different conflict analysis algorithm based on Algorithm 7 in [6] and removed some redundancies. I also added literal watching scheme and branching heuristics described in [7][11]. As a disadvantage of building upon this solver, the choice of data-structures was restricted, which had an effect on the overall efficiency of the solver.

The basic structure of CDPLL algorithms for Boolean SAT and many-valued SAT is the same. Decision and propagations are made until a falsified clause is found. Each decision literal increases the decision level. Every time a conflict is reached, a so-called no-good (clause representing an impossible assignment, derived from the "reason clauses" that lead to the conflict) is learned using a particular method (here resolution is used), typically aiming at first Unique Implication Point – the earliest propagation that causes the conflict. The learnt clause is then added to the clause database and the backtrack level is computed from it: upon backtrack the learned clause is unit, thus the propagation continues from the backtrack level. The learnt clause is implied by the original clause set, thus the addition doesn't change the problem semantically. Moreover, the most used VSIDS and Counter-based heuristics from Boolean SAT are easily adapted and already improve the search drastically.

```
Data: Problem in extended DIMACS format
Result: SAT / UNSAT
while checkSat() $\neq$ sat do
    if checkSat() = conflict then
        if level = 0 then
            return UNSAT
        end
        level = analyzeConflict();
        backtrack(level);
    end
    else if checkUnit() then
        propagate(unitLiteral);
    end
    else
        chooseLiteral();
        propagate(decisionLiteral);
    end
end
return SAT;
```

Algorithm 1: CDPLL

[11] Thanks to Irene Hiess for implementing the VSIDS heuristic.

The main difference between Boolean SAT CDPLL and many-valued SAT CDPLL lies in the propagation phase – when a positive literal is chosen, one also has to propagate the negative literals with the remaining values. This makes one choice more powerful and actually corresponding to many Boolean propagations. If all except one value on a variable are assigned, then the remaining positive literal is propagated (entail literal). Moreover, the generalized resolution is used, since there more contradicting combinations of literals than in Boolean SAT.

Structural Information As I mentioned before, the main advantage of the many-valued SAT in comparison to Boolean SAT is that the structural information is preserved in the many-valued formulation of the problem. When implementing a solver, one can exploit this feature in the following way: given the domain size of a variable and a number of appearances of a variable in a clause we have a threshold above which we know that the clause can still be satisfied, thus we don't have to visit such clauses. Here we don't have to know a specific value of a variable to be sure that the clauses where it appears are conflict-free. This can improve `checkSat()`, `checkUnit()` and `analyzeConflict()` procedures[12].

Conflict Analysis I implemented a resolution-based algorithm that computes the learned clause based on the first Unique Implication Point. The difference with Boolean SAT is that one also uses the entail clauses in the resolution: clauses stating that a variable should take at least one value from its domain. These are "lazy clauses" – they are invoked only when needed during conflict analysis and are not part of the clause database. This improves the efficiency of solving by making the formulation of the problem smaller.

Propagation Compared to SAT, there are more variations of propagation, simply because there are more types of choices possible. I follow the version which is closest to SAT: either positive or negative literal is propagated. However, it is also possible to choose or propagate several values at the same time. It is still unclear which decisions are more interesting: choosing a positive literal trims the search space considerably and leads to conflicts faster, especially in the case of 2-SAT problems. However, the learned clauses after such decisions are quite weak. Choosing a negative literal has less instant effect, since it removes only one value from the domain of a variable, but its propagating is faster and can lead to stronger propagation later after a clause is learned.

5 Future Work

5.1 Implementation

Data structures In order to take real advantage of the mentioned many-valued features, better data-structures are needed. For instance, in order to efficiently

[12] As currently implemented only the last effect is observable, in order to efficiently to perform this pre-check I am changing the data-structures.

perform propagation after many-valued choices (when not only positive literals are allowed to be chosen, but also literals with several possible values) one can use bitset representation of domains and then use the bitset operations which are very efficient. I am currently exploring this possibility. This way the watched literal scheme can be improved as described above. See [8] for more details.

Graph-based learning Now the clause-learning relies on resolution; however, there are other possibilities. In particular, there is a generalization of the Unique Implication Point method specific to many-valued setting that permits to learn stronger no-goods during the conflict analysis using the paths/cuts computations on the implication graph of the problem. However, computing such clauses is not linear as in our case [7]. But it may pay off since most of the time is spend on propagation, and it may be more beneficial to avoid increase in propagation rather than increase in time per conflict analysis.

Quality of learned clauses In SAT solving greedy learning scheme is used and emphasis is put on fast propagation, and not on quality of learned clauses. The currently winning solvers try to avoid this and concentrate on the quality of the learned clauses. They rely on the idea of *glue clauses* [4] – learned clauses that contain literals of only two levels. If such clauses are not removed and the solver aims at learning them the performance improves[13].

Restart strategies The experience of SAT shows that restarts are important in order to avoid bottle-necks in the search (also known as the heavy-tailedness phenomenon [10]). As we have seen from the n-queens example, the search even on easy problems can lead to wrong directions, thus such techniques should be implemented.

Heuristics Given the role of many-valued SAT as an intermediate between CSP and Boolean SAT, one could also use CSP heuristics for selecting a branching variable that proved to be effective [9].

Benchmarks To test the solver I developed some benchmarks in extended DIMACS format[14] `https://github.com/akinanop/mvl-solver/wiki/Benchmarks`, as well as used some existing ones. Namely, some graph coloring problems: `www-users.cs.york.ac.uk/~frisch/NB/`, `mat.gsia.cmu.edu/COLOR/instances.html#XXDSJ`, random binary CSP: `www.lirmm.fr/~bessiere/generator.html`, quasi-groups with holes: `www.cs.cornell.edu/gomes/gs-csgc.pdf`. There are few difficult problems that are not 2-SAT, thus it may be interesting to find more benchmarks of this type. However, we know that 2-SAT in many-valued setting is already NP-complete. This fact could be used to gain more efficiency by specializing the solver's data-structures and methods to 2-SAT problems.

[13] Take inspiration from `glucose`: `http://www.labri.fr/perso/lsimon/glucose/`.

[14] Thanks to Pavlo Myronov for implementing the graph coloring problems translator.

5.2 Theoretical investigation

One can explain why Boolean SAT solvers became efficient using the notion of a backdoor sets of variables [10]: a *backdoor set* is a set of variables of a propositional formula such that fixing the truth values of the variables in the backdoor set moves the formula into some polynomial-time decidable class. Current best heuristics guess these sets and then one solves polynomial sub-problems. Intuitively, a small backdoor set explains how a backtrack search can get "lucky" on certain runs: the backdoor variables are identified early on in the search and point in the right direction. It may be interesting to see whether the modelling and solving in many-valued SAT makes it easier to identify such sets earlier on some structured problems.

6 Conclusion

To summarize: in this project I developed a many-valued solver with several basic conflict-driven algorithms generalized from Boolean SAT. My experience shows that adapting the SAT algorithms to the many-valued setting can be worthwhile, given that the generalizations come naturally and don't require special theoretical effort. Provided the benefits of many-valued SAT solving mentioned in the literature and exemplified by the case study here, it seems like a fruitful direction of research.

References

1. Aloul, F.A., Ramani, A., Markov, I.L., Sakallah, K.A.: Solving difficult instances of boolean satisfiability in the presence of symmetry. IEEE Trans. on CAD of Integrated Circuits and Systems 22(9), 1117–1137 (2003), http://dx.doi.org/10.1109/TCAD.2003.816218
2. Ansótegui, C., Bofill, M., Manyà, F., Villaret, M.: Automated theorem provers for multiple-valued logics with satisfiability modulo theory solvers. Preprint submitted to Fuzzy Sets and Systems (2015)
3. Anstegui, C., Many, F.: Mapping problems with finite-domain variables into problems with boolean variables. In: In SAT 2004. pp. 1–15. Springer LNCS (2004)
4. Audemard, G., Simon, L.: Predicting learnt clauses quality in modern SAT solvers. In: IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009. pp. 399–404 (2009), http://ijcai.org/papers09/Papers/IJCAI09-074.pdf
5. Chepoi, V., Creignou, N., Hermann, M., Salzer, G.: The helly property and satisfiability of boolean formulas defined on set families. Eur. J. Comb. 31(2), 502–516 (2010), http://dx.doi.org/10.1016/j.ejc.2009.03.022
6. Jain, A.: Watched literals in a finite domain sat solver: Master thesis, University of Minnesota (2005), http://www.d.umn.edu/~jainx086/Thesis_Report.pdf
7. Jain, S., O'Mahony, E., Sellmann, M.: A complete multi-valued SAT solver. In: Principles and Practice of Constraint Programming - CP 2010 - 16th International Conference, CP 2010, St. Andrews, Scotland, UK, September 6-10, 2010. Proceedings. pp. 281–296 (2010), http://dx.doi.org/10.1007/978-3-642-15396-9_24

8. Liu, C., Kuehlmann, A., Moskewicz, M.W.: CAMA: A multi-valued satisfiability solver. In: 2003 International Conference on Computer-Aided Design, ICCAD 2003, San Jose, CA, USA, November 9-13, 2003. pp. 326–333 (2003), http://doi.ieeecomputersociety.org/10.1109/ICCAD.2003.1257732
9. Refalo, P.: Impact-based search strategies for constraint programming. In: Principles and Practice of Constraint Programming - CP 2004, 10th International Conference, CP 2004, Toronto, Canada, September 27 - October 1, 2004, Proceedings. pp. 557–571 (2004), http://dx.doi.org/10.1007/978-3-540-30201-8_41
10. Ryan Williams, C.G., Selman, B.: On the connections between backdoors, restarts, and heavy-tailedness in combinatorial search (2003), http://www.aladdin.cs.cmu.edu/papers/pdfs/y2003/sat7.pdf

Multi-Agent Epistemic Argumentation Logic

Chenwei Shi

Institute of Logic, Language and Computation, University of Amsterdam

Abstract. In this paper we build further on the recent work [12] on modelling an agent's beliefs directly in terms of the arguments that justify these beliefs. In particular, we extend the formal framework in [12] to a multi-agent setting. We analyze the relation between the agent's argumentation structure and her epistemic/doxastic state in this extended setting. Especially, we propose a way of defining the notion of belief based on the agent's argumentation structure. Moreover, we generalize the definition of belief to the definition of a group's distributed belief. And we show that the group's argumentation upon which the distributed belief is defined can be seen as a special form of a two-party argumentation. At last, we formalize the argumentation in the form of a two-player game to illustrate how the single agent's belief and the group's distributed belief are decided by the corresponding argumentation.

Keywords: Argumentation, Epistemic/Doxastic State, Modal Logic, Game

1 Introduction

Argumentation as a common human activity has been studied and analyzed in different fields: philosophy [14,9], game theory [8], and artificial intelligence [13,15], just to list a few. Different studies put emphasis on different aspects of "argumentation", for example, as a debating game between agents or as a reasoning process for belief formation and decision making. Despite of argumentation's multiple facets, the framework introduced in [4] succeeds in capturing some essential features of argumentation by highlighting the attack relation between arguments while abstracting away other details, as the way it formalizes the following example illustrates.

Example 1. In front of a vague picture of an animal, two people are arguing whether the animal in the picture is a bird:

- A: The animal in the picture has wings, so it is a bird (s_1);
- B: The animal looks like a bat, so it is not a bird (s_2);
- A: The animal does not only have wings but also have feathers, so it is a bird (s_3).

The formalization is given by the argumentation framework $\mathcal{AF} = \langle \mathcal{AR}, \twoheadleftarrow \rangle$ where $\mathcal{AR}$ is a set of arguments, and $\twoheadleftarrow$ is a binary relation on $\mathcal{AR}$. In the example, $\mathcal{AR} = \{s_1, s_2, s_3\}$ and $\twoheadleftarrow$ is an attack relation between these arguments

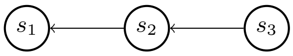


Fig. 1. Argumentation graph for Example 1

as shown in Figure 1. Note that the topic of the argumentation (whether it is a bird), the structure of each argument (the premises and the claims) and the agents involved in the argumentation are all ignored in the structure.

For different purposes, Dung's abstract argumentation framework is instantiated or extended. For example, in [10], it is instantiated with a general account of the structure of arguments and the nature of the attack relation; in [6] and [11], it is extended with a doxastic dimension for characterizing the notion of justified belief and modeling agents' beliefs about arguments in an argumentation respectively.

In this paper, we also add another dimension by introducing a set of possible worlds, in order to model each argument's claim (by assigning each argument node a set of possible worlds) and the topic of each argumentation (by labeling each attack relation with a set of possible worlds) in Dung's abstract argumentation framework (Section 2). However, different from [6] and [11], we define belief based on the argumentation structure rather than a doxastic relation in the added dimension. The way we relate belief to argumentation echoes Dung's idea on this issue in [4]:

> ...a statement is believable if it can be argued successfully against attacking arguments. In other words, whether or not a rational agent believes in a statement depends on whether or not the argument supporting this statement can be successfully defended against the counterarguments. ([4],p.323)

Moreover, we will show that this way of defining belief can be naturally generalized to a notion of group belief, i.e. distributed belief (Section 3.1). We can even take a further step by modeling a general form of argumentation between two parties which subsumes the argumentation structure needed for defining individual agent's belief and groups' distributed belief (Section 3.2). At last, we gamify this general form of argumentation (Section 3.3).

2 Multi-Agent Argumentation Logic

2.1 Extending the Argumentation Framework

In this section, we introduce our setting by extending the argumentation framework in [4].

Definition 1 (Argumentation-Support Frame (ASF)). *An argumentation-support frame is a structure* $\mathfrak{F} = \langle \mathcal{W}, \mathcal{AR}, \{\leftarrowtail^P_w\}^{P\subseteq\mathcal{W}}_{w\in\mathcal{W}}, \{f_w\}_{w\in\mathcal{W}}, g\rangle_{\mathcal{AG}}$ *where*

- $\mathcal{W} = \{u, v, w, \dots\}$ *is a non-empty set of possible worlds,* $\mathcal{AR} = \{s, t, \dots\}$ *is a non-empty set of arguments, and* $\mathcal{AG} = \{a, b, c, \dots\}$ *is a non-empty finite set of agents;*
- $\leftarrowtail^P_w \subseteq \mathcal{AR} \times \mathcal{AR}$ *assigns to each possible world* $w \in \mathcal{W}$ *an attack relation labelled by a subset* $P \subseteq \mathcal{W}$*;*

- *$f : \mathcal{AR} \times \mathcal{W} \to 2^{\mathcal{W}}$ assigns to each argument $s \in \mathcal{AR}$ in a possible world w a subset of $\mathcal{W}$ such that for any s and w, $f_w(s) \neq \varnothing$;*
- *$g : \mathcal{AG} \times \mathcal{W} \to 2^{\mathcal{AR}}$ assigns to each agent a in a possible world w a subset of $\mathcal{AR}$.*

By adding $\mathcal{W}$ and taking subsets of $\mathcal{W}$ as propositions, we distinguish between attack relations with respect to different topics. E.g. $s \leftarrowtail^P_w t$ expresses that argument s is attacked by t on the topic P in world w. Moreover, we make the claim supported by each argument explicit by introducing support function f. $f_w(s) \subseteq P$ expresses that argument s supports P. The other function g specifies the arguments each agent has in each possible world. Hence each agent in each possible world is assigned an argumentation structure which incorporates more details about the argumentation than an abstract argumentation framework:

$$\mathcal{AS}^{\mathfrak{F}}_{(a,w)} = \langle g_w(a), \{\leftarrowtail^P_{(a,w)} := \leftarrowtail^P_w \cap\, g_w(a) \times g_w(a)\}^{P \subseteq \mathcal{W}}, f_{(a,w)} := f_w \mid g_w(a)\rangle.$$

It is not hard to see that the argumentation structures for each agent in w are substructures of the following argumentation structure for possible world w which is neutral to any agent:

$$\mathcal{AS}^{\mathfrak{F}}_w = \langle \mathcal{AR}, \{\leftarrowtail^P_w\}^{P \subseteq \mathcal{W}}, f_w\rangle.$$

We will analyze the relation between the function g in the ASF and the agent's epistemic state which is usually modeled by an epistemic accessibility relation on possible worlds in the next section. Now let's turn to some conditions the argumentation-support frame should satisfy.[1] For notational simplicity, we write $\mathcal{W} - P$ as $\overline{P}$.

Definition 2. *Given any ASF, it should satisfy the following conditions:*

1. *if $s_1 \leftarrowtail^P_w s_2$, then $f(s_1) \subseteq P$ iff $f(s_2) \nsubseteq P$;*
2. *if $s_1 \leftarrowtail^P_w s_2$ and $f(s_1) \subseteq Q \subseteq P$, then $s_1 \leftarrowtail^Q s_2$ and if $s_2 \leftarrowtail^Q_w s_3$, then $s_2 \leftarrowtail^P s_3$.*

The first condition says that if s_2 attacks s_1 on P, then one of them must support P but not both of them. The second condition says that if s_2 attacks s_1 on P and s_1 supports a stronger claim Q which implies P, then s_2 also attacks s_1 on its claim Q. Moreover, if s_3 defends s_1 on its stronger claim Q, then s_3 also defends s_1 on P.

Remark 1. Note that the first condition implies that for any attack relation $\leftarrowtail^P_w$, no argument is controversial with respect to any argument. To put it more precisely, let $Att^P_w(s) = \{s_n \in \mathcal{AR} \mid \exists s_0, s_1, \ldots, s_n : \bigwedge_{i=0}^n s_i \leftarrowtail^P_w s_{i+1}$ where $s = s_0$ and n is an odd number $\}$ and $Def^P_w(s) = \{s_n \in \mathcal{AR} \mid \exists s_0, s_1, \ldots, s_n : \bigwedge_{i=0}^n s_i \leftarrowtail^P_w s_{i+1}$ where $s = s_0$ and $n \neq 0$ is an even number $\}$. We say that s_i is controversial with respect to s_j for the attack relation $\leftarrowtail^P_w$, if $s_i \in Att^P_w(s_j) \cap Def^P_w(s_j)$. The first condition implies that for any w, P and s, $Att^P_w(s) \cap Def^P_w(s) = \varnothing$.

[1] In [12], we impose stronger conditions. It is required that the arguments which attack an argument for P must support not P. In this paper, we only require that the arguments which attack an argument for P must not support P. In addition, we do not require that $\leftarrowtail^P_w$ should be the same as $\leftarrowtail^{\overline{P}}_w$.

2.2 Agent's Arguments and Knowledge

Function g in an ASF assigns to each agent in each possible world a set of arguments. If the argument s belongs to $g_w(a)$, then the argument s is available to the agent a in w. The way we construct the agent's argumentation structure implicitly presumes that once an argument belongs to $g_w(a)$, the agent knows not ony the argument itself and the claim it supports but also the attack relations between this argument and other arrguments belonging to $g_w(a)$. What is the difference between this way of modelling agents' knowledge of arguments and the way knowledge is modeled in epistemic logic (cf. [5])?

In epistemic logic, knowledge is defined based on an epistemic accessibility relation R_a for each agent a on the set of possible worlds. Generally, we assume that R_a is reflexive. The agent a knows a proposition P in w if and only if $R_a(w) \subseteq P$. Now given an ASF plus an epistemic accessibility relation for each agent $a \in \mathcal{AG}$, we can construct an argumentation structure for each agent based on the agent's epistemic accessibility relation R_a as follows:

$$\mathcal{EAS}^{\mathfrak{F}}_{(a,w)} = \langle \mathcal{AR}, \{\leftarrowtail^{P}_{(a,w)_e}\}^{P \subseteq \mathcal{W}}, f_{(a,w)_e} \rangle$$

where $\leftarrowtail^{P}_{(a,w)_e} = \bigcap_{v \in R_a(w)} \leftarrowtail^{P}_{v}$ and $f_{(a,w)_e}(s) = \bigcup_{v \in R_a(w)} f_v(s)$.

In $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$, once $s \in g_a(w)$, everything about the argument s is transparent to the agent (except the attack relation between it and other arguments which are not in $g_a(w)$). In this regard, we say that the agent knows argument s. Different from $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$, $\mathcal{EAS}^{\mathfrak{F}}_{(a,w)}$ only specifies the facts known by the agent about each argument. And not everything about the argument is known by the agent in $\mathcal{EAS}^{\mathfrak{F}}_{(a,w)}$. For example, there may be $s, t \in \mathcal{AR}$ such that $s \leftarrowtail^{P}_{w} t$ but $s \not\leftarrowtail^{P}_{(a,w)_e} t$ and $f_w(s) \subseteq P$ but $f_{(a,w)_e} \nsubseteq P$.

For simplicity, in this paper, we will take $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$ as the agent's epistemic argumentation structure. As we will see in the following sections, the simplification on this issue renders a uniform perspective on the relation between agents' argumentation structures and their beliefs possible. In the next section, we will focus on the single agent's belief, which is defined based on the agent's argumentation structure $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$.

2.3 Beliefs Supported by Arguments

We start by introducing some key notions of the abstract argumentation theory (see Figure 2) in [4] which can be characterized by the following two functions:

Definition 3. *Let $\mathfrak{F}$ be an ASF. The defense function $d^{P}_{(a,w)} : 2^{\mathcal{AR}} \to 2^{\mathcal{AR}}$ outputs a set of arguments defended by its input:*

$$d^{P}_{(a,w)}(X) = \{s \in \mathcal{AR} \mid \forall (s, s') \in \leftarrowtail^{P}_{(a,w)} \exists s'' \in X : (s', s'') \in \leftarrowtail^{P}_{(a,w)}\}.$$

And the neutrality function $n^{P}_{(a,w)} : 2^{\mathcal{AR}} \to 2^{\mathcal{AR}}$ outputs a set of arguments which is not attacked by its input:

$$n^{P}_{(a,w)}(X) = \{s \in \mathcal{AR} \mid \nexists\, s' \in X : s \leftarrowtail^{P}_{(a,w)} s'\}$$

X is conflict-free	iff	$X \subseteq n^{P}_{(a,w)}(X)$
X is self-acceptable	iff	$X \subseteq d^{P}_{(a,w)}(X)$
X is admissible	iff	$X \subseteq n^{P}_{(a,w)}(X)$ and $X \subseteq d^{P}_{(a,w)}(X)$
X is stable	iff	$X = n^{P}_{(a,w)}(X)$
X is a preferred extension	iff	X is a maximal admissible set
X is the grounded extension	iff	X is the least fixed point of $d^{P}_{(a,w)}$

Fig. 2. Some of the key notions of the abstract argumentation theory in [4]. For simplicity, we omit for each description its reference to the attack relation $\leftarrowtail^{P}_{(a,w)}$.

We say that an argument s for P is acceptable if there is an admissible set of arguments X such that $s \in X$, which means that s is defended by a set of arguments which is conflict free and can defend itself from other arguments' attack. If there is an acceptable argument s for P and there is no acceptable argument for $\overline{P}$ in the argumentation structure $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$, we say that agent a believes P in the possible world w.

Definition 4. *Given an ASF $\mathfrak{F}$, agent a* believes *P in the possible world w if there is an argument s and an admissible set of arguments X on P such that $s \in X$ and $f_w(s) \subseteq P$, but there is no argument t and an admissible set of arguments Y on $\overline{P}$ such that $t \in Y$ and $f_w(t) \subseteq \overline{P}$.*

Since each admissible set of arguments is a subset of a preferred extension and each preferred extension itself is admissible, the definition of belief actually requires that there is an argument s with $f_w(s) \subseteq P$ belonging to *the union of all preferred extensions* with respect to $\leftarrowtail^{P}_{(a,w)}$ but no argument t with $f_w(t) \subseteq \overline{P}$ belonging to *the union of all preferred extensions* with respect to $\leftarrowtail^{\overline{P}}_{(a,w)}$.

As we noted in Remark 1, given any $\mathcal{AS}^{\mathfrak{F}}_{w}$, it is uncontroversial in the sense of none of its arguments being controversial. Hence, for any agent a, $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$, as a substructure of $\mathcal{AS}^{\mathfrak{F}}_{w}$, is also uncontroversial. Thus according to Theorem 33 in [4],[2] it follows directly that

Proposition 1. *Given any structure $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$ and any $P \subseteq \mathcal{W}$, each of its preferred extension with respect to $\leftarrowtail^{P}_{(a,w)}$ is stable and its grounded extension with respect to $\leftarrowtail^{P}_{(a,w)}$ coincides with the intersection of all the preferred extensions with respect to $\leftarrowtail^{P}_{(a,w)}$.*

Therefore, there could be an alternative definition of belief which requires that there is an argument s with $f_w(s) \subseteq P$ belonging to *the intersection of all preferred extensions* with respect to $\leftarrowtail^{P}_{(a,w)}$ but no argument t with $f_w(t) \subseteq \overline{P}$ belonging to *the intersection of all preferred extensions* with respect to $\leftarrowtail^{\overline{P}}_{(a,w)}$.

[2] Theorem 33: Every uncontroversial argumentation framework's preferred extensions are stable and its grounded extension coincides with the intersection of all preferred extensions.

I.e. there is an argument s with $f_w(s) \subseteq P$ belonging to *the grounded extension* with respect to $\leftarrowtail^{P}_{(a,w)}$ but no argument t with $f_w(t) \subseteq \overline{P}$ belonging to *the grounded extension* with respect to $\leftarrowtail^{\overline{P}}_{(a,w)}$. Obviously, the notion of belief defined in this way implies the notion of belief defined in Definition 4. But in this paper, we will stick with the belief defined in Definition 4.

Moreover, we have the following proposition

Proposition 2. *Given an ASF $\mathfrak{F}$ and any argument s, there is an admissible set of arguments X with respect to $\leftarrowtail^{P}_{(a,w)}$ such that argument $s \in X$ if and only if $s \in Gfp.d^{P}_{(a,w)}$, where $Gfp.d^{P}_{(a,w)}$ is the greatest fixed point of the function $d^{P}_{(a,w)}$.*

Proof. We only sketch the proof here. Given the fact that if $X \subseteq d^{P}_{(a,w)}(X)$, then $X \subseteq Gfp.d^{P}_{(a,w)}$, the "only if" direction follows directly. For the "if" direction, assume that $s \in Gfp.d^{P}_{(a,w)}$, then take $X = Gfp.d^{P}_{(a,w)} \cap Def^{P}_{(a,w)}(s) \cup \{s\}$. We only need to check that $X \subseteq d^{P}_{(a,w)}(X)$ and $X \subseteq n^{P}_{(a,w)}(X)$.

So argument s is acceptable for agent a on P if and only if $s \in Gfp.d^{P}_{(a,w)}$. And the agent a believes (in the sense of Definition 4) P in the possible world w if and only if there is an argument s with $f_w(s) \subseteq P$ belonging to $Gfp.d^{P}_{(a,w)}$ but no argument t with $f_w(t) \subseteq \overline{P}$ belonging to $Gfp.d^{\overline{P}}_{(a,w)}$.

In the next section, we will present the language of the multi-agent argumentation logic and its truth conditions. In this language, we can express the notion of belief we define in Definition 4.

2.4 Syntax and Semantics

We start with the syntax of our logic:

Definition 5. *Let* Prop $= \{p, q, r, \dots\}$ *be a non-empty set of atomic propositions,* Anom $= \{\mathbf{s}, \mathbf{t}, \dots\}$ *be a non-empty set of argumentation nominals and* $\mathcal{AG} = \{a, b, c \dots\}$ *be a finite set of agents.* $\mathcal{L}$ *is the language generated by the following grammar:*

$$\alpha ::= \top \mid p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \boxbar_a \beta$$
$$\beta ::= \top \mid \mathbf{s} \mid \Box\alpha \mid \mathsf{Acc}^{\alpha}_{a} \mid \neg\beta \mid \beta \wedge \beta \mid [\leftarrowtail^{\alpha}]_a \beta$$

where $p \in$ Prop, $\mathbf{s} \in$ Anom *and* $a \in \mathcal{AG}$. *The duals of the operators are defined as usual, such as* $<\leftarrowtail^{\alpha}>_a$ *for* $\neg[\leftarrowtail^{\alpha}]_a\neg$.

The language is divided into two parts α and β. The α part is used to state facts about possible worlds, while β part is dedicated to the description of each argument. We will call formulas belong to α part (β part) of this language α formulas (β formulas). When there is no need to make distinction, φ is used to denote formulas in the whole language $\mathcal{L}$.

$\boxbar_a\beta$, as an α-formula, says that for all arguments of the agent a, β is the case. $\Box\alpha$ says that the current argument supports α and $[\leftarrowtail^{\alpha}]_a\beta$ says that for all arguments which are known by agent a and directly attack the current argument on α , β is the case. And Acc_a^{α} says that the current argument is acceptable to the agent a with respect to an argumentation on α.

The argument nominals give us the power of talking about arguments directly. For example, $\boxbar_a(\mathsf{s} \to \mathsf{Acc}_a^{\alpha})$ expresses that the argument s is an acceptable argument to the agent a.

In this language, the notion of belief defined in Definition 4 can be expressed as follows:

$$B_a\alpha := \diamondbar_a(\Box\alpha \wedge \mathsf{Acc}_a^{\alpha}) \wedge \neg \diamondbar_a (\Box\neg\alpha \wedge \mathsf{Acc}_a^{\neg\alpha}).$$

Note that we cannot have formulas as $B_a\beta$.

Remark 2. There are some interactions between α-formulas and β-formulas. For example, $\diamondbar_a \Box p$, which expresses that the agent a has an argument which supports p. However, not all interaction between these two formulas are allowed in the language. For example, the formulas like $\boxbar_a \boxbar_a \beta$, $\Box[\leftarrowtail^{\alpha}]_a\beta$, $[\leftarrowtail^{\alpha}]_a\alpha$ and $\mathsf{Acc}_a^{\Box\alpha}$. In the first formula, $\boxbar_a\beta$ expresses a fact about the possible worlds, so we cannot use it to describe arguments. In the second formula, $[\leftarrowtail^{\alpha}]_a\beta$ only describes a property of certain arguments, it is not a fact which can be supported.

Let $\mathfrak{M}$ be an argumentation-support model which is a triple $\langle \mathfrak{F}, \mathsf{n}, V\rangle$, where $\mathfrak{F}$ is an ASF, $V : \text{Prop} \to \mathcal{W}$ and $\mathsf{n} : \text{Anom} \to \mathcal{AR}$. Let $[\![\alpha]\!]_{\mathfrak{M}} = \{w \in \mathcal{W} \mid \mathfrak{M}, (w, s) \vDash \alpha\}$. We omit the subscript $\mathfrak{M}$ whenever possible. The truth of $\varphi \in \mathcal{L}$ is defined as follows:

Definition 6. *Given an argumentation-claim model $\mathfrak{M}$,*

$$\begin{array}{lll}
\mathfrak{M}, (w, s) \vDash \top & & \\
\mathfrak{M}, (w, s) \vDash p & \textbf{\textit{iff}} & w \in V(p) \\
\mathfrak{M}, (w, s) \vDash \mathsf{s} & \textbf{\textit{iff}} & \mathsf{n}(\mathsf{s}) = s \\
\mathfrak{M}, (w, s) \vDash \neg\varphi & \textbf{\textit{iff}} & \mathfrak{M}, (w, s) \nvDash \varphi \\
\mathfrak{M}, (w, s) \vDash \varphi \wedge \varphi' & \textbf{\textit{iff}} & \mathfrak{M}, (w, s) \vDash \varphi \textit{ and } \mathfrak{M}, (w, s) \vDash \varphi' \\
\mathfrak{M}, (w, s) \vDash \boxbar_a\beta & \textbf{\textit{iff}} & \textit{for any } s' \in g_w(a), \mathcal{M}, (w, s') \vDash \beta \\
\mathfrak{M}, (w, s) \vDash \Box\alpha & \textbf{\textit{iff}} & f_w(s) \subseteq [\![\alpha]\!] \\
\mathfrak{M}, (w, s) \vDash \mathsf{Acc}_a^{\alpha} & \textbf{\textit{iff}} & s \in Gfp.d_{(a,w)}^{[\![\alpha]\!]} \\
\mathfrak{M}, (w, s) \vDash [\leftarrowtail^{\alpha}]_a\beta & \textbf{\textit{iff}} & \textit{for any } s' \in g_w(a) \textit{ such that } s \leftarrowtail_{(a,w)}^{[\![\alpha]\!]} s', \mathfrak{M}, (w, s') \vDash \beta
\end{array}$$

The only case needs some extra attention is the truth condition of Acc_a^{α}, whose semantic meaning can be revealed by Proposition 2.

For the operator Acc_a^{α}, we have the following two properties:

Proposition 3. *Given any argumentation-supported model $\mathfrak{M}$, $\mathsf{Acc}_a^{\alpha} \to \mathsf{Acc}_a^{\alpha\vee\alpha'}$ is valid while $\mathsf{Acc}_a^{\alpha} \wedge \mathsf{Acc}_a^{\alpha'} \to \mathsf{Acc}_a^{\alpha\wedge\alpha'}$ is not.*

Proof. Take a pair (w, s) in the model $\mathcal{M}$ such that $\mathcal{M}, (w, s) \vDash \mathsf{Acc}_a^{\alpha}$. We need to prove that $\mathcal{M}, (w, s) \vDash \mathsf{Acc}_a^{\alpha\vee\alpha'}$. By Proposition 2, we only need to show that there is a set of argument X such that $s \in X \subseteq d^{[\![\alpha \vee \alpha']\!]}_{(a,w)}(X) \cap n^{[\![\alpha \vee \alpha']\!]}_{(a,w)}(X)$. Take $X = Gfp.d^{[\![\alpha]\!]}_{(a,w)} \cap Def^{[\![\alpha]\!]}_{(a,w)}(s) \cup \{s\}$. We first prove that $X \subseteq d^{[\![\alpha \vee \alpha']\!]}_{a}(X)$. Take any $t \in X$. If there is another argument t' such that $t \leftarrowtail^{[\![\alpha \vee \alpha']\!]}_{(a,w)} t'$. By condition 2 in Definition 2 and $f_w(t) \subseteq [\![\alpha]\!]$, it follows that $t \leftarrowtail^{[\![\alpha]\!]}_{(a,w)} t'$. Since $t \in X$ and $t \leftarrowtail^{[\![\alpha]\!]}_{(a,w)} t'$, there must be another argument $t'' \in Gfp.d^{[\![\alpha]\!]}_{(a,w)}$ such that $t' \leftarrowtail^{[\![\alpha]\!]}_{(a,w)} t''$. Together with condition 2 in Definition 2 and $f_w(t) \subseteq [\![\alpha]\!]$, it follows that $t' \leftarrowtail^{[\![\alpha \vee \alpha']\!]}_{(a,w)} t''$. Since $t'' \in Gfp.d^{[\![\alpha]\!]}_{(a,w)} \subseteq X$, $t \in d^{[\![\alpha \vee \alpha']\!]}_{a}(X)$. Next, we prove that $X \subseteq n^{\alpha \vee \alpha'}_{a}(X)$. Observe that for any $t \in X$, $f_w(t) \subseteq [\![\alpha]\!] \subseteq [\![\alpha \vee \alpha']\!]$. By condition 1 in Definition 2, it follows immediately that $X \subseteq n^{[\![\alpha \vee \alpha']\!]}_{a}(X)$.

It is not hard to come up with a counterexample against the validity of $\mathsf{Acc}^{\alpha}_{a} \wedge \mathsf{Acc}^{\alpha'}_{a} \to \mathsf{Acc}^{\alpha \wedge \alpha'}_{a}$. We leave it to readers.

It follows from this proposition that given any argumentation-supported model $\mathfrak{M}$, $B_a\alpha \to B_a(\alpha \vee \alpha')$ is valid. However, $B_a\alpha \wedge B_a\alpha \to B_a(\alpha \wedge \alpha')$ is not valid. From the condition that $f_w(s) \neq \varnothing$, it follows that $\neg B\bot$ is valid.

The way of defining single agent's belief based on an argumentation structure can be naturally generalized to characterize a group's distributed belief. In the next section, we will make the idea precise by generalizing the setting in this section.

3 Distributed Belief and Argumentation

Distributed knowledge is a standard notion in epistmic logic (cf. [5])). It is intended to characterize the knowledge a group of agents could get by combining all of its members' knowledge. Its semantic truth is based on the intersection of group members' sets of epistemically accessible worlds, which is taken as the the group's epistemically accessible worlds. Since knowledge implies truth, it is required that each agent's set of epistemically accessible worlds should include the actual world. So the intersection of them is always non-empty. However, when it comes to belief, if we still model it by a set of possible worlds, it is not reasonable any more to assume that the agent's doxcastically accessible worlds should include the actual world, since belief does not necessarily imply truth. So it is possible that the intersection of different agents' sets of doxastically accessible worlds is empty, which means that different agents' beliefs are inconsistent.

In this section, we will show that the argumentation structure provides with a way of reconciling the conflict between different agents' beliefs.

3.1 Distributed Belief

We first generalize the argumentation structure for a single agent $\mathcal{AS}^{\mathfrak{F}}_{(a,w)}$ to the argumentation structure for a group of agents. Let $g_w(D) := \bigcup_{a \in D} g_w(a)$.

$$\mathcal{AS}^{\mathfrak{F}}_{(D,w)} = \langle g_w(D), \{\leftarrowtail^{P}_{(D,w)} := \leftarrowtail^{P}_{w} \cap\, g_w(D) \times g_w(D)\}^{P \subseteq \mathcal{W}}, f_{(D,w)} := f_w \mid g_w(D)\rangle.$$

Correspondingly, we generalize the operator $\boxbar_a\beta$, $[\leftarrowtail^{\alpha}]_a\beta$ and Acc_a^{α} to $\boxbar_D\beta$, $[\leftarrowtail^{\alpha}]_D\beta$ and Acc_D^{α} respectively whose truth conditions are given as follows:

$$\mathfrak{M},(w,s)\vDash\boxbar_D\beta \quad \textbf{iff} \quad \text{for any } s'\in g_w(D),\mathcal{M},(w,s')\vDash\beta$$
$$\mathfrak{M},(w,s)\vDash\mathsf{Acc}_D^{\alpha} \quad \textbf{iff} \quad s\in Gfp.d_{(D,w)}^{[\![\alpha]\!]}$$
$$\mathfrak{M},(w,s)\vDash[\leftarrowtail^{\alpha}]_D\beta \ \textbf{iff} \quad \text{for any } s'\in g_w(D) \text{ such that } s\leftarrowtail_{(D,w)}^{[\![\alpha]\!]}s',\mathfrak{M},(w,s')\vDash\beta$$

It is not hard to see that $\boxbar_D\beta$ actually can be defined by $\bigwedge_{a\in D}\boxbar_a\beta$. However, this is not the case for $[\leftarrowtail^{\alpha}]_D\beta$ and Acc_D^{α}. Distributed belief can be defined as follows:

$$B_D\alpha := \diamondbar_D(\Box\alpha\wedge\mathsf{Acc}_D^{\alpha})\wedge\neg\diamondbar_D\,(\Box\neg\alpha\wedge\mathsf{Acc}_D^{\neg\alpha}).$$

Although the generalization made here is routine, the idea that distributed belief is decided by an argumentation within the group leads us to a more general perspective.

3.2 Argumentation between Agents

In this section, we show that the argumentation which decides a group's distributed belief as defined in the previous section is actually a special form of the argumentation between two parties, i.e. the proponent and the opponent.

First, we define an argumentation structure which represents the argumentation between two groups on certain topic Q:

Definition 7. *Given an ASF $\mathfrak{F}$, $\mathcal{AS}_{(D,E,w)}^{Q}=\langle\mathcal{AR}_{(D,E,w)}^{Q},\leftarrowtail_{(D,E,w)}^{Q},f_{(D,E,w)}^{Q}\rangle$ where*

- $\mathcal{AR}_{(D,E,w)}^{Q}=\{s\in g_w(D)\mid f_w(s)\subseteq Q\}\cup\{s\in g_w(E)\mid f_w(s)\nsubseteq Q\}$;
- $\leftarrowtail_{(D,E,w)}^{Q}:=\leftarrowtail_w^{P}\cap\mathcal{AR}_{(D,E,w)}^{Q}\times\mathcal{AR}_{(D,E,w)}^{Q}$;
- $f_{(D,E,w)}^{Q}:=f_w\mid\mathcal{AR}_{(D,E,w)}^{Q}$.

We stipulate that the group of agents taking the first position in the triple (D,E,w) is the proponent in the argumentation, while the group in the second position is the opponent. Since the argumentation is about Q, the proponent takes the burden of proving Q and the opponent needs to oppose by attacking the proponent's arguments. Hence, in this structure, the proponent only shows all its arguments which supports Q, while the opponent only shows all its arguments which does not support Q.

Note that if we take $D=E$ in $\mathcal{AS}_{(D,E,w)}^{Q}$, then $\leftarrowtail_{(D,E,w)}^{Q}=\leftarrowtail_{(D,w)}^{Q}$ where $\leftarrowtail_{(D,w)}^{Q}$ is defined in $\mathcal{AS}_{(D,w)}^{\mathfrak{F}}$. Hence $\mathcal{AS}_{(D,w)}^{\mathfrak{F}}$ actually presents the argumentation by a group of agents D itself on everything in which D does not hide any arguments.

$\boxbar_D\beta$, $[\leftarrowtail^{\alpha}]_D\beta$ and Acc_D^{α} in the language can be correspondingly generalized to $\boxbar_{D,E}^{\alpha}\beta$, $[\leftarrowtail^{\alpha}]_{D,E}\beta$ and $\mathsf{Acc}_{D,E}^{\alpha}$ respectively:

$$\mathfrak{M},(w,s)\vDash\boxbar_{D,E}^{\alpha}\beta \quad \textbf{iff} \quad \text{for any } s'\in\mathcal{AR}_{(D,E,w)}^{[\![\alpha]\!]},\mathcal{M},(w,s')\vDash\beta;$$
$$\mathfrak{M},(w,s)\vDash\mathsf{Acc}_{D,E}^{\alpha} \quad \textbf{iff} \quad s\in Gfp.d_{(D,E,w)}^{[\![\alpha]\!]};$$
$$\mathfrak{M},(w,s)\vDash[\leftarrowtail^{\alpha}]_{D,E}\beta \ \textbf{iff} \quad \text{for any } s'\in\mathcal{AR}_{(D,E,w)}^{[\![\alpha]\!]} \text{ such that } s\leftarrowtail_{(D,w)}^{[\![\alpha]\!]}s',\mathfrak{M},(w,s')\vDash\beta.$$

The function $d^Q_{(D,E,w)}$ is defined in the same pattern as $d^Q_{(a,w)}$ except that the attack relation $\leftarrowtail^Q_{(a,w)}$ in the definition now becomes $\leftarrowtail^Q_{(D,E,w)}$.

Note that $\boxplus_D\beta$, $[\leftarrowtail^\alpha]_D\beta$ and Acc^α_D can be translated into $\boxplus^\top_{D,D}\beta$, $[\leftarrowtail^\alpha]_{D,D}\beta$ and $\mathsf{Acc}^\alpha_{D,D}$.

The argumentation structure represents the argumentation in a static way. In the next section, we will try to define for each argumentation structure a corresponding extensive argumentation game.

3.3 Argumentation Game

Argumentation is a process rather than a static structure. In this section, we present the argumentation structure in a form of extensive game, which helps illustrate the idea that belief formation is essentially a process of argumentation.

Definition 8 (Two-player argumentation game). *Given an ASF $\mathfrak{F}$ and an argumentation structure $\mathcal{AS}^Q_{(D,E,w)}$, a two-player argumentation game on Q denoted by $\mathcal{TG}^Q_{(D,E,w)}$ consists of (we omit the subscript and superscript for the notations introduced below when it is clear from the context.)*

- *two players $D, E \subseteq \mathcal{AG}$ with D being the proponent of Q and E being the opponent of Q*
- *the arsenal of D as a proponent of Q is $\mathcal{PAR} = \{s \in g_w(D) \mid f_w(s) \subseteq Q\}$ and the arsenal of E as an opponent of P is $\mathcal{OAR} = \{s \in g_w(E) \mid f_w(s) \nsubseteq Q\}$;*
- *a Turn function such that*
  - *$Turn(0) \in \mathcal{PAR}$;*
  - *if $m = 2n+1$, then $Turn(m) \subseteq \{t \in \mathcal{OAR}_w \mid (s,t) \subseteq \leftarrowtail^Q_{(D,E,w)}$ where $s \in Turn(m-1)\}$.*
  - *if $m = 2n > 0$, then $Turn(m) \subseteq \{t \in \mathcal{PAR} \mid (s,t) \in \leftarrowtail^Q_{(D,E,w)}$ where $s \in Turn(m-1)\}$.*

Note that we allow in the game that the players can present more than one argument in each move except the first move.

In this game the winning conditions for the proponent and the opponent are a little different, since the burden of proof is on the proponent. The winning condition for D is that $Turn(0) \neq \varnothing$ and for any $m = 2n$, there is $s \in Turn(m-1)$ such that $t \in Turn(m)$ such that $s \leftarrowtail^Q_{(D,E,w)} t$ (the $\forall\exists$ - pattern). The winning condition for E is that for any $m = 2n+1$, there is $s \in Turn(m-1)$ such that there is $t \in Turn(m)$ with $s \leftarrowtail^Q_{(D,E,w)} t$ (the $\exists\exists$ - pattern).

Given the winning conditions for each player, we can define the winning strategies for each player. Let $Pick_D : 2^{\mathcal{OAR}} \to 2^{\mathcal{PAR}}$ ($Pick_E : 2^{\mathcal{PAR}} \to 2^{\mathcal{OAR}}$) be a strategy for the proponent (opponent). $Pick_D$ is a winning strategy for D if for any function $Turn$ such that $Turn(0) = Pick_D(\varnothing)$ and $Turn(2n) = Pick_D(Turn(2n-1))$ with $n > 0$, it satisfies the wining condition for D. $Pick_E$ is a winning strategy for E if for any function $Turn$ such that $Turn(2n+1) = Pick_E(Turn(2n))$, it satisfies the winning condition for E.

Proposition 4. *Given an ASF $\mathfrak{F}$, an argumentation structure $\mathcal{AS}^{Q}_{(D,E,w)}$ and its corresponding two-player argumentation game $\mathcal{TG}^{Q}_{(D,E,w)}$, $Pick_D$ is a winning strategy for the proponent D if and only if $Pick(\varnothing)_D \in GFP.d^{Q}_{(D,E,w)}$.*

Proof. We only sketch the proof here.

For the "only if" direction, let $Pick_D$ be the winning strategy for D. We only need to show that $s = Pick_D(\varnothing) \in GFP.d^{Q}_{(D,E,w)}$. we prove this by contraposition. Suppose that $s \notin GFP.d^{Q}_{(D,E,w)}$. So there is no $X \subseteq \mathcal{AR}^{Q}_{(D,E,w)}$ such that $s \in X \subseteq d^{Q}_{(D,E,w)}(X)$. It implies that for any subset of $Def^{Q}_{(D,E,w)}(s)$, denoted by SD, $SD \cup \{s\} \nsubseteq d^{Q}_{(D,E,w)}(SD \cup \{s\})$, which means there must be $t \in Att^{Q}_{(D,E,w)}(s)$ such that t attacks $SD \cup \{s\}$ and there is no argument $t' \in SD \cup \{s\}$ with $t \leftarrowtail^{Q}_{(d,E,w)} t'$. Now we construct a function $Turn$ such that $Turn(0) = Pick_D(\varnothing)$, $Turn(2n+1) = \{t \in \mathcal{OAR}_w \mid (s,t) \subseteq \leftarrowtail^{Q}_{(D,E,w)}$ where $s \in Turn(m-1)\}$ and $Turn(2n) = Pick_D(Turn(2n-1))$. So no matter what $\bigcup_n Turn(2n) \subseteq Def^{Q}_{(D,E,w)}(s)$ is, there is $t \in Att^{Q}_{(D,E,w)}(s)$ such that t attacks $\bigcup_n Turn(2n)$. So there must be an odd number i such that $t \in Turn(i)$. However, there is no $t' \in \bigcup_n Turn(2n)$ such that $t \leftarrowtail^{Q}_{(d,E,w)} t'$. So $Turn$ does not satisfy the winning condition for D. So $Pick_D$ is not a winning strategy for D.

For the "if" direction, we just construct a strategy $Pick_D$ with $Pick_D(\varnothing) \in GFP.d^{Q}_{(D,E,w)}$ such that for any $X \in 2^{\mathcal{OAR}}$, $Pick_D(X) = \{t \in \mathcal{PAR} \mid (s,t) \in \leftarrowtail^{Q}_{(D,E,w)}$ where $s \in X\} \cap GFP.d^{Q}_{(D,E,w)}$. So we only need to show that $Pick_D$ is a winning strategy for D.

4 Conclusion and Future Work

In this paper, we start with an extension of Dung's argumentation framework in which we can formalize the notion of argument-supported belief. And this notion of belief can be naturally generalized to a notion of distributed belief which is defined based on the group's argumentation. Moreover, we show that the argumentation upon which the group's distributed belief is based is a special form of a two-player argumentation game. And the last proposition in this paper reveals the relation between the winning strategy in an argumentation game and the acceptability of arguments (in line with the use of greatest fixed point in [3] for defining solution concepts in strategic game-theoretic contexts). Alongside the analysis, we devise a logic to express all these notions. It can be taken as a preliminary attempt on "merging dynamic logics of information flow with concrete models of argumentation" as van Benthem suggests in [2].

Therefore, it will be a natural follow-up to take into account the change of the agents' arguments and its influence on the agents' belief. And as a special notion of belief, its relation to other notions of belief is also an interesting topic, for example, probabilistic notion of belief [7] and the notion of belief defined in the plausibility model [1]. In addition, axiomatization, decidibility and other properties of the proposed logic needs a further study.

References

1. Baltag, A., Smets, S.: A qualitative theory of dynamic interactive belief revision. Texts in logic and games 3, 9–58 (2008)
2. van Benthem, J.: One logician's perspective on argumentation. Cogency 1, 13–26 (2009)
3. van Benthem, J.: Logic in Games. MIT press (2014)
4. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial intelligence 77, 321–357 (1995)
5. Fagin, R., Moses, Y., Vardi, M.Y., Halpern, J.Y.: Reasoning about knowledge. MIT press (2003)
6. Grossi, D., van der Hoek, W.: Justified beliefs by justified arguments. In: Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning (2014)
7. Halpern, J.Y.: Reasoning about Uncertainty. MIT press (2003)
8. Jacob, G., Rubinstein, A.: Debates and decisions: On a rationale of argumentation rules. Games and Economic Behavior 36, 158–173 (2001)
9. Perelman, C.: The new rhetoric. Springer (1971)
10. Prakken, H.: An abstract framework for argumentation with structured arguments. Argument and Computation 1(2), 93–124 (2010)
11. Schwarzentruber, F., Vesic, S., Rienstra, T.: Building an epistemic logic for argumentation. In: Proceedings of JELIA'12. pp. 359–371 (2012)
12. Shi, C., Sonja: Beliefs supported by arguments, in proceedings of CLAR, to appear.
13. Simari, G.R., Loui, R.P.: A mathematical treatment of defeasible reasoning and its implementation. Artificial intelligence 52, 125–157 (1991)
14. Toulmin, S.E.: The uses of argument. Cambridge University Press (2003)
15. Vreeswijk, G.A.: Abstract argumentation systems. Artificial intelligence 90, 225–279 (1997)

Logic of Closeness Revision
Challenging relations in social networks[*]

Anthia Solaki, Zoi Terzopoulou, and Bonan Zhao

ILLC, University of Amsterdam

Abstract. In social epistemology and dynamic epistemic logic (DEL), the study of belief revision and opinion dynamics in social networks has recently gained increasing attention. Our social contacts affect the way we form our opinions about the world. However, in many real life situations we can also observe the dual effect: people's opinions may also play a role in the evolution of the network's structure. In this paper, we present a complete logic that models the dynamical changes in the agents' network relations with respect to opinion exchange. We make use of a 2×2 coordination game, the "discussion game". We first focus on the simplified cases where issues are equally weighted, agents never change opinion on them, and just modify their network relations accordingly. Next, we introduce different weights on issues in order to express agents' priorities. Finally, we discuss an extension of our model that can capture more refined schemata of human interaction.

1 Introduction

After Claire met Frank, she found that this young man shared similar opinions and attitudes towards most things with her. Love therefore grew between the two. Sometimes they had quarrels, and big opinion differences almost led them to breaking up, but the same goal of achieving power always united them and gave them strength. This is the story of House of Cards in three sentences. In this paper, we will develop logical tools based on a game-theoretical framework in order to answer the following question: *how* does opinion exchange affect our closeness with our social contacts?

According to the standard approach of belief revision in DEL, agents are continuously under the influence of their network-neighbors and modify their opinions in the view of the social norm. However, as far as empirical evidence is concerned, these attempts seem defective. Consider a "stubborn" agent: she stands on firm to her opinions; once she interacts with someone, she reshapes her relationship with him. Naturally, agreement is viewed as a positive boost for a

[*] We thank Paolo Galeazzi for his valuable intuitions and his time, and Alexandru Baltag for his comments. Zoi's Terzopoulou gratitude is also owned to the Onassis foundation for financing her studies and her research work at the University of Amsterdam. The paper was initially developed during a project course in the ILLC, offered by Johan van Benthem and Fenrong Liu.

relationship, while conflict is a burden. Belief revision, as presented for example in Baltag et al. 2015, does not account for similar scenarios. It is precisely this gap that this paper wishes to bridge.

We should emphasize that what follows is not a marginal, case-study of some peculiar agents. For instance, any agent can be treated as a stubborn agent at a certain context, in the sense that some issues may trigger non-negotiable opinions which cannot be subject of change merely due to the diffusion of a fashion in a network. Agreement and disagreement among agents in a network can characterize the network's structure and evolution. In Section 2, a coordination game is introduced, in order to capture opinion differences between a pair of agents, on a given issue. The model of closeness revision and its update are grounded on this "discussion game". The fact that different issues attract more attention than others, depending on the agents that interact, is also incorporated in the framework of this paper. In Section 3, we present a complete dynamic logic, which combines probabilistic and qualitative logics, in order to reason about the dynamics in a network. To conclude, realistic human interactions prescribe that agents behave stubbornly or not, depending on their social environment; this variation of the model is discussed in Section 4.

2 The Model of Closeness Revision

Our setting consists of finite networks of agents and *closeness* relations among them. Closeness relations are weighted (ranging from -1, reflecting hate, to 1, reflecting total closeness) and not necessarily symmetric.[1] Next, we assume that there is a countable set of issues according to which agents re-evaluate their relations in the network. The motivating idea reads as follows: although two agents may already know each other's opinion before discussing an issue, the discussion on it can be iterated, and their relation may be re-evaluated. For instance, I may know that we disagree on the political proposals of the Republican Party of the US, but re-asserting our divergent opinions makes me upset and subsequently decreases our closeness. We define a mathematical model which represents the revision of the closeness relation according to the *measure of agreement* between agents, and include it into a logical system. The definition of *the model of closeness revision* will serve as the building block for all the illustrations that will follow.

Definition 1. *A model of closeness revision is a tuple*

$$M = \langle A, C, I, O \rangle$$

where A is a set of agents, $C : A \times A \to [-1, 1]$, with $C(a, b)$ interpreted as the closeness relation between agents a and b, I is a set of issues, and $O : A \times I \to [0, 1]$, with $O(a, i)$ interpreted as the opinion of agent a on issue i.

[1] We drop symmetry, even if it is very commonly used in other works on the topic (Liu, Seligman, and Girard 2014, Baltag et al. 2015, Christoff, Hansen, and Proietti 2014), as not-necessarily symmetric relations seem to better capture cases where feelings between agents are not mutual.

Note that we allow $C(a, a)$, which can be read as “closeness to oneself”, to take any value in $[-1, 1]$. We will further assume that closeness to oneself is unaltered, not affected by any issue-induced revision. The latter assumption reflects the idea of one’s stable relationship to herself, and modelling different personality types will not concern us for the purposes of this paper.
We denote **C** the class of models of closeness revision.

2.1 The Discussion Game

In the model of closeness revision, any agent a holds an opinion towards any given issue i, denoted as $O(a, i)$. When a pair of agents a, b discusses about the issue i, what is of our interest is the degree on which $O(a, i)$ and $O(b, i)$ deviate from each other. Assume for simplicity that every agent announces her opinion truthfully during the discussion. Opinions’ divergence transforms agents’ closeness afterwards.

Discussion entails interaction between agents.[2] The satisfaction that two agents a and b gain from a discussion on the issue i progresses with respect to their agreement on i. Thus, we can view the discussion on i as a coordination game, and interpret agents’ pure opinions towards the issue i as their strategies in the *discussion game*. Let AG and DG be the pure strategies that express agreement and disagreement on i respectively.

i	AG	DG
AG	1;1	-1;-1
DG	-1;-1	1;1

Then, take $(O(a, i)AG, (1-O(a, i))DG)$, with $O(a, i) \in [0, 1]$ to be the mixed strategy of agent a over agreement and disagreement on i. Consider, for example, that i is the issue “proposals of the Republican party”. $O(a, i) = 0$ tells us that agent a does not like such proposals at all, having probability zero to agree on them. $O(a, i) = \frac{1}{2}$ suggests that agent a is equally expected to agree or disagree with the proposals and yields out an indifferent state of opinion. $O(a, i) = 1$ indicates that agent a fully agrees with the proposals.

		$O(b, i)$	$1 - O(b, i)$
	i	AG	DG
$O(a, i)$	AG	1;1	-1;-1
$1 - O(a, i)$	DG	-1;-1	1;1

[2] It is common to consider discussion as a strategic situation where cooperation is preferred. Approaches that involve competition between agents who discuss will not be considered in this paper, but are suggested for further research.

The expected utility of agents a and b in this game is interpreted as their *measure of agreement* on i[3]: $V^i(O_a, O_b) := O_a O_b - O_a(1 - O_b) - (1 - O_a)O_b + (1 - O_a)(1 - O_b)$. To simplify the formulation we will write $V^i(a, b)$ instead of $V^i(O_a, O_b)$. We will see that cooperation between agents in the discussion game will increase their closeness after playing it.

2.2 Model Update

Once we have the model, we want to update it to capture the dynamics in social interaction. We make use of the 2×2 discussion game to define our update function.

Definition 2. *The update of the model $M = \langle A, C, I, O\rangle$ over an issue $i \in I$ is the model $M^i = \langle A, C^i, I, O\rangle$, where C^i is given by the following formula.*

$$C^i(a, b) = \frac{C(a, b) + V^i(a, b)}{2}$$

and $C^i(a, a) = C(a, a)$.

The updating function captures the impact of the measure of agreement on the agents' closeness. We will motivate the use of this function with a number of real life examples.[4]

Example 1. **Discussion with an indifferent agent.**
Consider agents Alice (a) and Bob (b) who argue on the proposals of the Republican party. In particular, the issue is: the party's proposal on military expenditure (m). Suppose that the agents are very close to each other, that is, $C(a, b) = C(b, a) = C(a, a) = C(b, b) = 1$. However, agent a strongly supports the party's policy of military procurement whereas agent b is indifferent. Formally: $O(a, m) = 1$, $O(b, m) = 0.5$. According to our model, the measure of agreement is $V^m(a, b) = 0$, and subsequently the agents' closeness will be $C^m(a, b) = C^m(b, a) = 0.5$.

The above example reflects the scenario where an agent is indifferent on a thorny political issue, and this can indeed be proven to be harmful for her relationship with a strong supporter of this issue.

Example 2. **Hostile agents get closer when they agree...**
On the contrary, let agents Claire (c) and Dan (d) be such that $C(c, c) = C(d, d) = 1$ and $C(c, d) = -0.6$, $C(d, c) = -1$. They, too, engage in a political conversation over the issue m with both agreeing on it, having $O(c, m) = 1$ and $O(d, m) = 0.9$. According to our model, the measure of agreement is $V^m(c, d) = 0.8$ and their revised closeness is $C^m(c, d) = 0.1$ and $C^m(d, c) = -0.1$.

[3] Where for simplicity we write O_a instead of $O(a, i)$ and O_b instead of $O(b, i)$.

[4] We should note that in this framework –and throughout the paper in general– we assume that agents perform their revisions simultaneously. Therefore, an agent's judgment is only affected by the previous-stage data and not by the possible shifts other agents make in the current stage.

Therefore, although the agents were initially hostile, their strong political agreement brought them closer and made them provisionally more tolerant towards each other.

Example 3. **...But not for long.**
Next, assume that Claire and Dan keep discussing about Republican proposals, introducing the issue of corporate tax (t). Suppose that $O(c,t) = 0.2$ and $O(d,t) = 1$. Then, their measure of agreement is $V^t(c,d) = -0.6$ and now their revised closeness will be $C^t(c,d) = -0.25$ and $C^t(d,c) = -0.35$.

Overall, the two agents' divergent opinions on the second political issue decreased the shaky closeness they acquired after their first agreement. Triggered by this example and talking in general terms, predictions and insights on the long-term behavior of a network can be accommodated once the particular model is employed.[5]

Example 4. **The order of the discussed issues matters.**
Finally, suppose that Claire and Dan discussed the same issues presented before, but discussion on the corporate tax preceded the one on military spending. Our model prescribes that, after the first round of discussion, the revised closeness will be $C^t(c,d) = -0.6, C^t(d,c) = -0.8$, and after the second round of discussion, $C^m(c,d) = 0.1$ and $C^m(d,c) = 0$.

Example 4 illustrates the following

Proposition 1. *Closeness revision is order-dependent.*

Indeed, the fluctuations of a relationship can be reasonably accounted in terms of the alternations of agreement and disagreement over time. Specifically, in real life scenarios, the impact that the discussion of an issue can have on a relationship does not only depend on the issue itself, but also on the context in which the discussion takes place (the issues that have been discussed before, etc.)

Of course, once we have the discussion games in our toolbox, the updating attempt is not unique. Depending on the scenario that is modeled, additional constraints can be established and the updated closeness might also be calculated in a different manner. In Section 4, we propose a refinement of our model update.

2.3 Weighted Issues

It is also reasonable to consider the priority that an agent gives to a specific issue. We expect that the more important an issue is for an agent a, the more it affects a's relations. We represent agent a's priority over issues by adding a weighting function into our model, $W_a : I \to [0,1]$, where $W_a(i) = 0$ reflects no importance, and $W_a(i) = 1$ reflects the highest priority. Therefore, the agents'

[5] General results on networks' evolution using the framework of this paper are open for further investigation.

payoffs in the discussion game may differ depending on the weights. For example, if $O(a,i) = O(b,i) = 1$ but $W_a(i) > W_b(i)$, we expect that agent a will get higher subjective utility by agreeing with b, because the issue i is more important to her. Overall, the payoffs of the discussion game express the "amount of satisfaction" for each agent after the discussion.

		$O(b,i)$	$1 - O(b,i)$
	i	AG	DG
$O(a,i)$	AG	W_a;W_b	-W_a;-W_b
$1 - O(a,i)$	DG	-W_a;-W_b	W_a;W_b

A model of Closeness Revision with weighted issues and its update can be defined accordingly.

Definition 3. *A model of Closeness Revision with weights is a tuple* $M = \langle A, C, I, O, (W_a)_{a \in A} \rangle$.

Let us now consider the following example.

Example 5. **Discussion between agents with different priorities.**
Alice (a) is very close to Ben (b), she supports the Republican policy on military expenditure and this also constitutes one of her top priorities. On the contrary, Ben disagrees with it, but he places military concerns low in his agenda. Formally, take: $C(a,b) = C(b,a) = C(a,a) = C(b,b) = 1$, $O(a,m) = 1$, $O(b,m) = 0$, $W_a(m) = 1$ and $W_b(m) = 0$. The measure of agreement is $V^m(a,b) = -1$. According to the model with weighted issues: $C^m(a,b) = 0$ and $C^m(b,a) = 0.5$.

In other words, following the update, Alice becomes utterly distant to Ben, due to Ben's disagreement on an issue that is so essential for her. Yet Ben, despite slightly shifting away from Alice, still regards her relatively close.

Hopefully the above example convinced the reader that adding weights in the model is a step closer to the idea of imitating real life scenarios.

3 The Logic of Closeness Revision

In this section, we present a complete dynamic logic to capture the notions that have been described so far. The logic is based on the model in Definition 1 and its update in Definition 2, and is inspired by techniques used in logics for reasoning about probability (Fagin, Halpern, and Megiddo 1990; Van Benthem, Gerbrandy, and Kooi 2009).

3.1 Syntax and Semantics

Definition 4. *Let A be a finite set and I be a countable set.*
The set $\mathcal{T}$ of terms contains the sets of constants $\{C_{ab} : a, b \in A\}$, $\{O_{ai} : a \in A, i \in I\}$ and $\{V_{abi} : a, b \in A, i \in I\}$.

For $q_1, \ldots, q_n \in \mathcal{T}$ and $a_1, \ldots, a_k, c \in \mathbf{Z}$, the set $\mathcal{A}$ contains atoms of the form $\alpha_1 q_1 + \ldots + \alpha_k q_k \geq c$.
Let $\Phi := \mathcal{T} \cup \mathcal{A}$ be the set of all primitive propositions.
The Language of Closeness Revision $\mathcal{L}_{CR}$ is defined as follows:

$$p \in \Phi \mid \neg\phi \mid \phi \wedge \phi \mid [i]\phi$$

Intuitively, the set $\mathcal{T}$ indicates facts about the agents' closeness and opinions, whereas the set $\mathcal{A}$ suggests numerical inequalities between the values representing closeness and opinions. The $[i]$ modality is interpreted as in standard dynamic epistemic logic (Van Ditmarsch, Der Hoek, and Kooi 2007; Van Benthem and Liu 2007): we evaluate $[i]\phi$ as true "today" if and only if ϕ is true "tomorrow" after the revision induced by issue i.
The symbols $\vee$, $\rightarrow$, $-$, $\leq$, $>$, $<$, $=$ are defined in the usual way. For example the formula $q = c$ stands as abbreviation for $(q \geq c) \wedge ((-1)q \geq -c)$. Moreover, a formula with rational numbers such as $q > \frac{1}{5}$ can be expressed by $5q > 1$. So, we can always allow rational numbers in our formulas as abbreviations for the formula that can be obtained by clearing the denominators.

Definition 5. *Let $M = \langle A, C, I, O \rangle$ be a model of closeness revision as defined in Definition 1. The interpretation q^M of terms q in M is defined as follows: $C^M_{ab} := C(a, b)$, $O^M_{ai} := O(a, i)$ and $V^M_{abi} := V^i(a, b)$.*

Given a model $M = \langle A, C, I, O \rangle$, the truth clauses for $\mathcal{L}_{CR}$ are the following.

- $M \models \alpha_1 q_1 + \ldots + \alpha_k q_k \geq c$ iff $\alpha_1 q^M_1 + \ldots + \alpha_k q^M_k \geq c$
- $M \models \neg\phi$ iff $M \not\models \phi$
- $M \models \phi \wedge \psi$ iff $M \models \phi$ and $M \models \psi$
- $M \models [i]\phi$ iff $M^i \models \phi$

Abbreviations We introduce the following abbreviations $t^{[i]}$ in order to capture (in the logical language) the values of terms $t \in \mathcal{T}$ after the revision with issue i, according to Definition 2.

- $O^{[i]}_{aj} := O_{aj}$, for any $j \in I$
- $V^{[i]}_{abj} := V_{abj}$, for any $j \in I$
- $C^{[i]}_{ab} := \frac{1}{2}C_{ab} + \frac{1}{2}V_{abi}$, for $a \neq b$
- $C^{[i]}_{aa} := C_{aa}$

3.2 Complete Axiomatization

The system L_{CR} that we present divides nicely into three parts, which deal respectively with propositional reasoning, reasoning about linear inequalities and reasoning about dynamics. We obtain a complete axiomatization of the logic for the models of closeness revision and their updates, by using the standard technique of reduction laws from DEL (Van Ditmarsch, Der Hoek, and Kooi 2007; Blackburn, De Rijke, and Venema 2002).

Definition 6. *The following axiom system is sound and complete with respect to the class of models* $\mathbf{C}$.

All instances of valid formulas for propositional logic	*Prop*
All instances of valid formulas for linear inequalities	*Ineq*
$0 \leq O_{ai} \leq 1$	*Bound O*
$-1 \leq C_{abi} \leq 1$	*Bound C*
$0 \leq V_{abi} \leq 1$	*Bound V*
$O_{ai} = v \wedge O_{bi} = w \rightarrow V_{abi} = u$ *for all* $v, w, u \in [0,1]$ *s.t.* $vw - v(1-w) - (1-v)w + (1-v)(1-w) = u$	*Cor.* O, V
$[i]((\sum_{m=1}^{k} a_m q_m) \geq c) \leftrightarrow ((\sum_{m=1}^{k} a_m q_m^{[i]}) \geq c)$ *for all* $k \in \mathbf{N}$	*Red.Ax.Ineq*
$[i](\phi \wedge \psi) \leftrightarrow [i]\phi \wedge [i]\psi$	*Red.Ax.* $\wedge$
$[i]\neg\phi \leftrightarrow \neg[i]\phi$	*Red.Ax.* $\neg$
From ϕ *and* $\phi \rightarrow \psi$, *infer* ψ	*Modus Ponens*

The static part of the logic consists of the axioms of propositional logic Prop, the axiom Ineq, the Bounding axioms for opinion, closeness and measure of agreement, the correlation axiom between O and V, and the rule of Modus Ponens. In order to deal with the dynamic part of the logic, we need rules which reduce formulas that contain the $[i]$ modality to formulas without it. This is possible, as all the information required to determine the updated model M^i is present in the model M before the update. The reduction laws are trivial in all cases apart from those involving atoms of the form $\alpha_1 q_1 + \ldots + \alpha_k q_k \geq c$, for $a_1, \ldots, a_k, c \in \mathbf{Z}$. By making use of the abbreviations presented before, the axiom Red.Ax.Ineq encodes the numerical changes on the terms' values.

Theorem 1 (Completeness). *For any* $\phi \in \mathcal{L}_{CR}$, *we have that*

$$\vDash_{\mathbf{C}} \phi \quad \textit{iff} \quad \vdash_{L_{CR}} \phi.$$

Proof. Soundness: Let $M = \langle A, C, I, O \rangle$ be an arbitrary model, with $a \in A$ and $i \in I$. The axiom Ineq is easily checked to be true on M, as the updates are on atoms, so sophisticated checks whether we can stay inside the language of linear inequalities are not required. Then, the formulas $0 \leq O_{ai} \leq 1$, $-1 \leq C_{abi} \leq 1$ and $0 \leq V_{abi} \leq 1$ are satisfied, by the way the model is defined. Definition 1

combined with the definition of the measure of agreement can also verify that the Cor. O, V axiom is satisfied. Soundness of Red.Ax.$\wedge$ and Red.Ax.$\neg$ can be shown by induction on the structure of the formulas.
Completeness: Fagin, Halpern, and Megiddo 1990 provide us with a complete axiomatization of all valid formulas about linear inequalities. The axioms Bound O, Bound C and Bound V guarantee that the numerical bounds on opinion, closeness and measure of agreement are provable in our system. Moreover, the axiom Cor. O, V ensures that the correlation between agent's opinions and their measure of agreement, as defined in our framework, is provable.
Finally, we can translate the dynamic part of the language into its static part using the reduction laws given above. Then, the proof goes in the standard way (Van Ditmarsch, Der Hoek, and Kooi 2007). □

3.3 Safe Friends, Future Friends and Dangerous Issues

We now present some supplementary definitions that support putting the logical framework in practical context.

Given a certain friendship threshold θ_F we say that:
Agent b is agent a's *friend* (F_{ab}) whenever $C(a, b)$ is above θ_F. Therefore $M \vDash F_{ab}$ iff $M \vDash C_{ab} \geq \theta_F$. We call a and b friends whenever $M \vDash F_{ab}$ and $M \vDash F_{ba}$.
Agent b is a's *safe friend* for issue i (SF^i_{ab}) whenever b is a's friend and the revision induced by i cannot break this friendship, that is $M \vDash SF^i_{ab}$ iff $M \vDash F_{ab} \wedge [i]C_{ab} \geq \theta_F$. We call a and b safe friends whenever a is b's safe friend and b is a's safe friend.
Agent a is a *future friend* of agent b given issue i (FF^i_{ab}) whenever b is not a's friend, yet after the revision induced by i friendship is established. Formally, $M \vDash FF^i_{ab}$ iff $M \vDash \neg F_{ab} \wedge [i]C_{ab} \geq \theta_F$. We call a and b future friends whenever $M \vDash FF^i_{ab}$ and $M \vDash FF^i_{ba}$.
An issue i is *dangerous* for a's friendship with b (D^i_{ab}) whenever F_{ab} is true before the update induced by i, but not after, namely $M \vDash D^i_{ab}$ iff $M \vDash F_{ab} \wedge [i]\neg F_{ab}$.

When we combine the previous framework with the notions above, we can observe that further validities hold.

- $[i]F_{ab} \leftrightarrow SF^i_{ab} \vee FF^i_{ab}$: b is a's friend after discussing i iff b is a's safe friend for issue i or i establishes a future friendship.
- $SF^i_{ab} \vee FF^i_{ab} \vee [i]\neg F_{ab}$: given issue i, if b is neither safe nor future friend with a, then b is not going to be a's friend after discussing i.
- $F_{ab} \wedge V_{abi} \geq C_{ab} \rightarrow [i]F_{ab}$: if b is already a's friend and the value of agreement on issue i is higher than the closeness value of a and b, then discussing i will not break the friendship.

4 Model with Updating Threshold

In this section, a variation of the model of closeness revision is provided, with modifications required to deal with potential challenges. We will enrich the model update of Definition 2 adding a further threshold condition on the update function.

Even if so far we claimed that all agents *can be considered* stubborn with respect to certain issues, it is still plausible to argue that not all agents *are* always stubborn. Agreement and disagreement affect people's relationships in different degrees. In this part of the paper we consider that only agents who have enough social closeness can afford being stubborn.[6] If being stubborn can lead an agent to be socially isolated, this agent is more conservative about updating her closeness with her social contacts. In other words, if updating closeness will result in social isolation, agents do not perform the update.

Definition 7. *A model of closeness revision with threshold is a tuple*

$$M_\theta = \langle A, C, I, O, \theta \rangle$$

where A , C, I, O *as in Definition 1 and* $\theta \in [0,1]$ *is a threshold for revising closeness.*

The threshold for revising closeness represents the level of closeness that an agent does not feel comfortable to go below. Consequently, agents will behave stubbornly and keep revising only when their closeness level is above the threshold θ.

Definition 8. *The update of the model of closeness revision with threshold* $M_\theta = \langle A, C, I, O, \theta \rangle$ *over an issue* $i \in I$ *is the model* $M_\theta^i = \langle A, C^i, I, O, \theta \rangle$, *where* C^i *is given by the following formula.*

$$C^i(a,b) = \begin{cases} \frac{C(a,b)+V^i(a,b)}{2} & \textit{if } \frac{\sum_{d\in A} C(a,d)}{|A|} \geq \theta \textit{ or } C(a,b) < V^i(a,b), \textit{ and } a \neq b \\ C(a,b) & \textit{else} \end{cases}$$

The condition $\frac{\sum_{d\in A} C(a,d)}{|A|} \geq \theta$ or $C(a,b) < V^i(a,b)$ says that for agent a to revise her closeness with agent b over the issue i, she needs to be safe enough to be stubborn (expressed by the formula $\frac{\sum_{d\in A} C(a,d)}{|A|} \geq \theta$), or if she revises, she will increase the value of her "social closeness" (reflected by the formula $C(a,b) < V^i(a,b)$ as follows: $\frac{C(a,b)+V^i(a,b)}{2} > C(a,b) \Leftrightarrow C(a,b) + V^i(a,b) > 2C(a,b) \Leftrightarrow C(a,b) < V^i(a,b)$).

Example 6. **Who behaves stubbornly after all?**
Suppose that Claire and Dan are the only agents in the network, and they discuss the Republican proposal on military expenditure (m). Their closeness relations are represented by the values: $C(c,c) = C(d,d) = 1$, $C(c,d) = 0.5$ and $C(d,c) = 0.1$. Suppose that their opinions are $O(c,m) = 1$ and $O(d,m) = 0.5$,

[6] Sufficiency of social closeness will be captured by a threshold condition. For the purposes of this paper, the threshold will be uniform for all the agents. However, different thresholds can be added to express different agents' tendency to stubbornness.

that is, Claire strongly agrees on the issue, while Dan is indifferent. It follows that their measure of agreement on m is $V^m(c,d) = 0$. Let the threshold for revising closeness be $\theta = 0.6$. According to the model of closeness revision with threshold:

- For Claire, the condition for closeness revision is satisfied, as $\frac{C(c,c)+C(c,d)}{2} = 0.75 > 0.6$. This means that Claire feels confident enough to behave stubbornly. Therefore, she will revise her closeness with Dan, having $C^m(c,d) = 0.25$ after the discussion.
- For Dan, however, the condition for closeness revision is not satisfied, as $\frac{C(d,d)+C(d,c)}{2} = 0.55 < 0.6$ and $C(d,c) = 0.1 > 0 = V^m(c,d)$. This means that Dan does not have enough social closeness, so he does not revise his relationships. Therefore, his closeness with Claire remains the same $C^m(d,c) = C(d,c) = 0.1$ after the discussion.

Overall, this scenario demonstrates how two different agents may behave stubbornly or not, according to their social closeness at the moment of a discussion.

5 Conclusion and Further Research

To sum up, in this paper we use a game-theoretical approach to build a dynamical model that is able to represent the interactive revision of both agents' opinions and agents' relations in a social network. Measures of agreement and disagreement are used to define the revision dynamics of these two main dimensions considered here. Specifically, the interaction between agents who discuss is expressed by the discussion game that captures agreement and disagreement on a specific issue under consideration. We finally introduce a sound and complete axiom system for models of closeness revision.

A first limitation of our framework concerns an implicit assumption: the willingness of agents to cooperate, reflected by the discussion game. One could reasonably argue that her closeness with someone may increase not only in situations of agreement, but also in cases of *constructive disagreement*. This is an intriguing issue to reflect on, even though a counter-argument would support that the number of people in a network who would appreciate a disagreement as fruitful is so small that becomes insignificant. Still in the direction of our design choices, an objection can be raised regarding the difference between the *quantitative* and the *qualitative* side of an opinion's expression. The function O captures the former, while the latter is ignored. In the presented framework, an agent merely announces the content of her opinion, that is the degree on which she agrees with the discussed issue. However, real life examples suggest that the *strength* of opinions plays a principal role in human interactions, too. Defining opinion as a twofold notion, with both a quantitative and a qualitative part, and modifying the revising conditions accordingly, would be a natural and appealing extension of our model. Some other questions that deserve further investigation are: Firstly, concerning the game-theoretical part: How can we strengthen the connection of our framework with games? General results on network studies can provide more insights into the topic of networks' dynamical changes triggered by agents' discussions. Moreover, techniques from evolutionary game theory could be proven to be useful in analyzing how profitable human interactions of certain kind are for a society, or for a social network. Secondly, on the logical part: How can we extend the Logic of Closeness Revision to capture more refined schemata of interaction, as for instance the one presented in Section 4? In real life scenarios, it is also possible to observe the combined action of two different dynamics between connected agents: opinions affecting relations and relations affecting opinions. Agents in different contexts may behave stubbornly or revise their opinions instead. Furthermore, the model of closeness revision does not take into account the level of information that agents have about the opinions of the others in their network. So, in which way could our logic evolve, if we add epistemic -indistinguishability- relations for agents? To conclude with a philosophically oriented concern: Is it possible, in logical and mathematical terms, to value the future of such a complicated concept, as a human relationship?

References

[Bal+15] Alexandru Baltag et al. "Dynamic Epistemic Logics of Diffusion and Prediction in Social Networks". In: *Draftpaper, April* (2015).

[BDV02] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal Logic: Graph. Darst.* Vol. 53. Cambridge University Press, 2002.

[CHP14] Zoé Christoff, Jens Ulrik Hansen, and Carlo Proietti. "Reflecting on social influence in networks". In: *Information Dynamics in Artificial Societies Workshop.* 2014.

[FHM90] Ronald Fagin, Joseph Y Halpern, and Nimrod Megiddo. "A logic for reasoning about probabilities". In: *Information and computation* 87.1 (1990), pp. 78–128.

[LSG14] Fenrong Liu, Jeremy Seligman, and Patrick Girard. "Logical dynamics of belief change in the community". In: *Synthese* 191.11 (2014), pp. 2403–2431.

[VDK07] Hans Van Ditmarsch, Wiebe van Der Hoek, and Barteld Kooi. *Dynamic epistemic logic.* Vol. 337. Springer Science & Business Media, 2007.

[VGK09] Johan Van Benthem, Jelle Gerbrandy, and Barteld Kooi. "Dynamic update with probabilities". In: *Studia Logica* 93.1 (2009), pp. 67–96.

[VL07] Johan Van Benthem and Fenrong Liu. "Dynamic logic of preference upgrade". In: *Journal of Applied Non-Classical Logics* 17.2 (2007), pp. 157–182.

A Qualitative Analysis of Kernel Extension for Higher Order Proof Checking

Shuai Wang*

INRIA Rocquencourt, France
ILLC, University of Amsterdam, The Netherlands
Email: shuai.wang@student.uva.nl

Abstract. For the sake of reliability, the kernels of Interactive Theorem Provers (ITPs) are kept relatively small in general. On top of the kernel, additional symbols and inference rules are defined. This paper presents the first qualitative analysis how kernel extension reduces the size of proofs and impact proof checking.

Keywords: HOL Light, proof checking, kernel extension

1 Introduction

Higher order logic is also known as simple type theory. It is an extension of simply typed λ-calculus with additional axioms and inference rules [4]. Interactive Theorem Provers (ITPs) of higher order logic have been playing an important role in formal mathematics, software verification and hardware verification. However, ITPs may have bugs and may lead to errors in proofs generated while not being apparent within the proof systems themselves. Also, proofs nowadays can be huge, making it difficult or even impossible to check by hand. For example, the Kepler Conjecture project took a team of scientists several years with many ITPs involved [5]. The demand of reliability of such ITPs makes proof checking necessary, especially by proof checkers independent from the ITPs. Taking advantage of the similarity of the logic and design between some ITPs, OpenTheory [9] has developed a standard format for serialising proofs [9]. One way to verify these proofs (also known as proof articles) is to export them to the OpenTheory format followed by the proof checking process by Dedukti [11].
The correctness of an ITP depends on its kernel where basic symbols and inference rules are defined [6, 10]. On top of the kernel, more symbols and corresponding inference rules are defined. The kernel of HOL Light takes equality as its only logical (term) symbol to keep the size of its kernel minimal. Some dependency analyses of the symbols of the HOL Light system show that, aside from equality, there is also much dependency on implication and universal quantification. In contrast, HOL4 takes conjunction, disjunction, implication, existential quantification and so on as primitive symbols. This paper presents

*The author was supported by the MPRI-INRIA scholarship during this project.

HOLALA, an alternative version of HOL Light with a kernel extension of additional symbols and inference rules. More specifically, implication and universal quantification were taken primitive. This paper presents a first experimental work on qualitative measurement of the impact of kernel extension with a concentration on proof checking efficiency.
This paper is organised as follows: Chapter 2 explains the kernel of HOL Light and Chapter 3 illustrates the design of HOLALA by extending the kernel of HOL Light. Following that is the update of Holide and Dedukti as well as proof checking and evaluation in Chapter 4.

2 HOL Light

Higher order logic is also known as simple type theory. It is a logic on top of simply typed λ-calculus with additional axioms and inference rules [4]. The type of a term is either an individual, a boolean type or a function type. A term is either a constant, a variable (e.g. x), an abstraction (e.g. $\lambda x.x$) or a well-typed application (e.g. $(\lambda x.x)y$). The notation $x : \iota$ means that the term x is of type ι. Types are sometimes omitted for simplicity of representation.

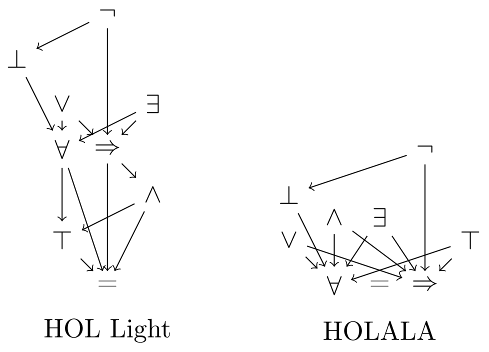
type variables	α, β
type operators	p
types	$A, B ::= \alpha \,|\, p(A_1, \ldots, A_n)$
term variables	x, y
term constants	c
terms	$M, N ::= x \,|\, \lambda x : A.M \,|\, MN \,|\, c$

HOL Light [7] is an open source interactive theorem prover for higher order logic. Its logic is an extension of Church's Simple Type Theory [2] with polymorphic type [7]. The kernel of HOL Light is an OCaml file where terms, types, symbols and inference rules are defined. Symbols and inference rules in the kernel are considered primitive. On top of the kernel, additional symbols are introduced and inference rules are derived. The kernel of HOL Light has only one primitive logical (term) symbol, the equality $(=)$[1]. The equality is of polymorphic type [8] and plays three roles in HOL Light: definition, equivalence and bi-implication.

3 Kernel Extension

On top of the kernel, more logic connectives and constants are introduced. Figure 1 illustrates the dependency of these symbols based on their definition

[1] $s = t$ is a conventional concrete syntax for $((=)st)$.



HOL Light HOLALA

Fig. 1: Dependency Analysis

	HOL Light	HOLALA
$=$	primitive	primitive
$\Rightarrow$	$\lambda pq.p \wedge q \Leftrightarrow p$	primitive
$\forall$	$\lambda p.(p = \lambda x.\top)$	primitive
$\Leftrightarrow$	$=$	$=$
$\exists$	$\lambda p \forall q(\forall x.px \Rightarrow q) \Rightarrow q$	$\lambda p \forall q(\forall x.px \Rightarrow q) \Rightarrow q$
$\top$	$\lambda p.p = \lambda p.p$	$\forall x.(x \Rightarrow x)$
$\bot$	$\forall p.p$	$\forall p.p$
$\wedge$	$\lambda pq.(\lambda f.fpq) = (\lambda f.f\top\top)$	$\lambda pq.(\forall x.(p \Rightarrow ((q \Rightarrow x) \Rightarrow x)))$
$\vee$	$\lambda pq.\forall r.(p \Rightarrow r) \Rightarrow ((q \Rightarrow r) \Rightarrow r)$	$\lambda pq.\forall r.(p \Rightarrow r) \Rightarrow ((q \Rightarrow r) \Rightarrow r)$
$\neg$	$\lambda p.p \Rightarrow \bot$	$\lambda p.p \Rightarrow \bot$

Table 1: Primitive and Axiomatic Definitions of Connectives and Constants Comparison

as in Table 1. For example, the definition of ∃ depends on that of ⇒. Note that equality is in fact used when introducing every symbol but the graph omits such arrows for the sake of simplicity. It can be observed that logical connectives have much dependency on implication and universal quantification as well. This leads to the idea of introducing them as primitive symbols to reduce the depth of dependency and shorten proofs without changing proof scripts.

The kernel also includes ten primitive inference rules as in Table 2 (with types eliminated to keep the table small). On the base of the ten primitive inference rules, we introduce derived inference rules. The correctness of the proofs largely depends on the correctness of the kernel [6, 10].

The kernel of HOL-style ITPs are generally kept small for the sake of reliability. A kernel provides primitive types, core inference rules and constants and various safe definitional mechanisms. If correctly implemented (assume the correctness of the meta-language), the soundness of the ITP is guaranteed. Kernels vary from small ones (e.g. HOL Light and HOL Zero) to larger ones

Structural	$\dfrac{}{\{A\} \vdash A}\ \mathit{ASSUME}$
λ Calulus	$\dfrac{\Gamma \vdash A = B}{\Gamma \vdash \lambda x.A = \lambda x.B}\ \mathit{ABS}$ $\dfrac{}{(\lambda x.A)x = A}\ \mathit{BETA}$
Instantiation	$\dfrac{\Gamma[x_1,\ldots,x_n] \vdash A[x_1,\ldots,x_n]}{\Gamma[t_1,\ldots,t_n] \vdash A[t_1,\ldots,t_n]}\ \mathit{INST}$ $\dfrac{\Gamma[\alpha_1,\ldots,\alpha_n] \vdash A[\alpha_1,\ldots,\alpha_n]}{\Gamma[\gamma_1,\ldots,\gamma_n] \vdash A[\gamma_1,\ldots,\gamma_n]}\ \mathit{INST\_TYPE}$
Bi-implication	$\dfrac{\Gamma \vdash A = B \qquad \Delta \vdash A}{\Gamma \cup \Delta \vdash B}\ \mathit{EQ\_MP}$ $\dfrac{\Gamma \vdash A \qquad \Delta \vdash B}{(\Gamma \setminus \{B\}) \cup \{\Delta \setminus \{A\}) \vdash A = B}\ \mathit{DEDUCT\_ANTISYM\_RULE}$
Equality	$\dfrac{}{\vdash A = A}\ \mathit{REFL}$ $\dfrac{\Gamma \vdash A = B \qquad \Delta \vdash C = D}{\Gamma \cup \Delta \vdash A(C) = B(D)}\ \mathit{MK\_COMB}$ $\dfrac{\Gamma \vdash A = B \qquad \Delta \vdash B = C}{\Gamma \cup \Delta \vdash A = C}\ \mathit{TRANS}$

Table 2: Primitive Inference Rules of HOL Light [7]

(e.g. HOL4). The more constants and inference rules taken primitive in the kernel, the harder it is to guarantee the soundness of the system. Although it is known to the HOL community that correctly expanding a kernel would lead to some efficiency gains, there is no qualitative measurement of this benefit. This paper shows how the extension of kernels would reduce the depth of dependency, leading to a reduction of the size of proofs and a speedup of proof checking without the loss of reliability. We introduce HOLALA, a modified version of (OpenTheory) HOL Light[2] where the kernel consists of more logic symbols and their corresponding inference rules. Different from HOL Light which takes equality as the only primitive symbol, HOLALA has an extended the HOL Light kernel with universal quantification and implication and their associated introduction and elimination rules (*MP*, *GEN*, *DISCH* and *SPEC*). This was achieved by adding the universal quantifier and implication symbol

[2] (OpenTheory) HOL Light is HOL Light equipped with proof recording methods and exports proofs into proof packages, namely the article files. (OpenTheory) HOL Light also generates the standard library of the OpenTheory Repository. We refer to (OpenTheory) HOL Light as HOL Light in the rest of this paper for short (despite the differences in some detailed proofs in each systems and other aspects).

to the kernel[3]. In addition, HOLALA also modified the definition of truth (⊤), and conjunction (∧), making as many definitions of logic symbols as possible dependent on the universal quantifier and implication instead. The definition of symbols of HOLALA in comparison with HOL Light is shown in Table 1. To summarise, Figure 1 shows a comparison of the dependency of symbols in HOL Light and HOLALA. As a consequence, some derived inference rules were reproved. An immediate benefit of such changes is that the derived inference rules directly depending on inference rules of implication and universal quantification were shortened. For example, the conjunction introduction rule is expanded to 31 inference steps instead of 55 while recording. Similarly, the disjunction introduction rule takes 21 inference steps instead of 156. For this reason, proofs are expected to be shorter.

$$\frac{\Gamma \vdash A \Rightarrow B \qquad \Delta \vdash A}{\Gamma \cup \Delta \vdash B}\, MP$$

$$\frac{\Gamma \vdash A[c/x]}{\Gamma \vdash \forall x A}\, GEN \text{ if x is not free in } \Gamma$$

$$\frac{\Gamma \vdash B}{\Gamma \setminus \{A\} \vdash A \Rightarrow B}\, DISCH$$

$$\frac{\Gamma \vdash \forall x A}{\Gamma \vdash A[t/x]}\, SPEC$$

Although such changes lead to the reduction of proof size, users would lose the original definitions of the ∀ and ⇒. To fix proofs explicitly involving these two definitions, the definitions of the ∀ and ⇒ are proved as theorems after the introduction of the axiom of extensionality.

4 Proof Checking and Evaluation

4.1 Extending Holide and Dedukti

In this project we employ Dedukti as the proof checker to verify the proofs generated by HOLALA. Cousineau and Dowek showed that Higher Order Logic can be embedded in the $\lambda\Pi$-calculus Modulo as well as other Pure Type Systems (PTS) [3]. This laid the foundation of Dedukti [11], a universal proof checker. On top of Dedukti, Holide[1] was developed to transform proofs from a proof repository, namely the OpenTheory Repository [9], to Dedukti. Following the extension of the logic kernel of HOL Light, there are some necessary changes to the existing translation of HOL Light's logic into Dedukti to accommodate this larger kernel. To deal with this update, the declaration of the universal quantifier and the implication together with their elimination and introduction inference rules were added to Holide as well as the input to Dedukti. The quantified terms would be translated as follows, with the notation of translation follows from the notation of [1]:

[3] Note that, similar to equality, universal quantification is also of polymorphic type.

$| \to | = imp$
$|(\forall_A)| = forall|A|$
$|\forall(M : A)| = forall|A||M|$
$|M \to N| = imp|M||N|$, where
imp: $term\,bool \to term\,bool\ \to term\,bool$
forall : $\Pi\alpha : type \to term(arr\,\alpha\,bool) \to term\,bool$

To translate the additional inference rules, four constants MP, $DISCH$, GEN and $SPEC$ were introduced as below:

MP: $\Pi p : term\,\mathrm{bool}.\Pi q : term\ bool.proof(imp\,p\,q) \to proof p \to \mathrm{proof}\,q$
DISCH: $\Pi p : term\ bool.\Pi q : term\ bool.proof p \to proof\ q \to proof(imp\,p\,q)$
GEN: $\Pi\alpha : type.\Pi p' : (term\,\alpha \to term\,bool).\Pi x : term\,\alpha.proof(p'\,x) \to proof(forall\,\lambda\,x.p'x)$
SPEC: $\Pi\alpha : type.\Pi t : (term\,\alpha \to term\,bool).\Pi u : term\,\alpha.proof(forall\,\alpha\,t) \to proof(t\,u)$

The translation of corresponding inference rules were added to Holide:

$\left| \dfrac{\Gamma \vdash A \Rightarrow B \qquad \Delta \vdash A}{\Gamma \cup \Delta \vdash B}\ MP \right| = \mathrm{MP}|A||B||\mathcal{D}_1||\mathcal{D}_2|$, where $\mathcal{D}_1$ and $\mathcal{D}_2$ are the proofs of $A \Rightarrow B$ and A respectively.
$\left| \dfrac{\Gamma \vdash A[c/x]}{\Gamma \vdash \forall x A}\ \textit{GEN, if } x \textit{ is not free in } \Gamma \right| = \mathrm{GEN}|A||c'||\mathcal{D}'|$, where $c' = \lambda x : ||A||.|c|$, $\mathcal{D}$ is a proof of $A[c/x]$ and $\mathcal{D}' = \lambda x : ||A||.|\mathcal{D}|$
$\left| \dfrac{\Gamma \vdash B}{\Gamma \setminus \{A\} \vdash A \Rightarrow B}\ DISCH \right| = \mathrm{DISCH}|A||B||\mathcal{D}'||\mathcal{D}|$, where $\mathcal{D}'$ is a proof of A and $\mathcal{D}$ is a proof of B
$\left| \dfrac{\Gamma \vdash \forall x A}{\Gamma \vdash A[t/x]}\ SPEC \right| = \mathrm{SPEC}\,|A|t'|u||\mathcal{D}|$, where $t' = \lambda x : ||A||.|t|$ and $\mathcal{D}$ is a proof of B.

4.2 Evaluation

A way to compare proof size is to consider the size of the article files. To reduce the effect of syntax formatting and white-space, all the article files and Dedukti files from both systems are compressed by *gzip*. The size of both article files and Dedukti files scale down considerably after compression. Here we take the (OpenTheory) HOL Light's standard theory library for evaluation. As shown in Table 3, the average size of the article files of HOLALA is around 64.36% that of OpenTheory. This leads to an improvement of 41.81% in translation time. The size of Dedukti files were reduced to about 64.92% with an acceleration of 38.04% for proof checking.

	Size of Proof Files (KB)	Translation Time (s)
HOL Light	5,376	55.98
HOLALA	3,460	32.57
Comparison	Reduced to 64.36%	Improved by 41.81%

	Size of Dedukti Files (KB)	Proof Checking Time (s)
HOL Light	16,092	30.75
HOLALA	10,448	19.05
Comparison	Reduced to 64.92%	Improved by 38.04%

Table 3: Comparison of Translation and Proof Checking

5 Conclusion and Discussion

An optimal design of a HOL kernel comes in various point of views: size and complexity, reasoning speed and memory efficiency, consideration of proof checking, etc. This paper presented HOLALA, a variant of HOL Light with an extended kernel by introducing implication and universal quantification. We provided the first qualitative measurement of the reduction the proof size and the speed up proof checking. The size of proofs of HOLALA reduced to 64.36% on average, leading to an improvement of a speed-up of 38.04% for proof checking. It also worth noting that ITPs are usually developed without much concern about the size of proofs and the complexity of proof checking. This paper attempted to bring theorem proving and proof checking closer with an emphasis on the efficiency of proof checking. While OpenTheory grounds proofs to a minimal representation using a variant of HOL Light, this work shows the potential to ground proofs to a more efficient representation corresponding to a bigger (or the maximal) kernel instead. This work could be further completed by introducing conjunction and disjunction, truth and false, existential quantifier and more to the kernel. Another possible future work is to import proofs to (a variant of) HOL4 and export proofs out for further efficiency testing. Following this line, some further comparative experiments may be conducted between different extended kernels and the best efficiency payoff compared to its size.

Bibliography

[1] Ali Assaf and Guillaume Burel. Translating HOL to dedukti. In *Proceedings Fourth Workshop on Proof eXchange for Theorem Proving, PxTP 2015, Berlin, Germany, August 2-3, 2015.*, pages 74–88, 2015.

[2] Henk Barendregt, Wil Dekkers, and Richard Statman. *Lambda calculus with types*. Cambridge University Press, 2013.

[3] Denis Cousineau and Gilles Dowek. Embedding pure type systems in the lambda-pi-calculus modulo. In *Typed lambda calculi and applications*, pages 102–117. Springer, 2007.

[4] William M Farmer. The seven virtues of simple type theory. *Journal of Applied Logic*, 6(3):267–286, 2008.

[5] Thomas C Hales, John Harrison, Sean McLaughlin, Tobias Nipkow, Steven Obua, and Roland Zumkeller. A revision of the proof of the kepler conjecture. In *The Kepler Conjecture*, pages 341–376. Springer, 2011.

[6] John Harrison. Towards self-verification of hol light. In *Automated Reasoning*, pages 177–191. Springer, 2006.

[7] John Harrison. HOL Light: An overview. In *Theorem Proving in Higher Order Logics*, pages 60–66. Springer, 2009.

[8] Leon Henkin. A theory of prepositional types. *Fundamenta Mathematicae*, 3(52):323–344, 1963.

[9] Joe Hurd. The opentheory standard theory library. In *NASA Formal Methods*, pages 177–191. Springer, 2011.

[10] Magnus O Myreen, Scott Owens, and Ramana Kumar. Steps towards verified implementations of hol light. In *Interactive Theorem Proving*, pages 490–495. Springer, 2013.

[11] Ronan Saillard. Dedukti: a universal proof checker. In *Foundation of Mathematics for Computer-Aided Formalization Workshop*, 2013.

Exhaustivity Effect of Focus-Particles in German it-Clefts: Empirical and Corpus-based Insights

Anna-Christina Boell

CRC Text Structures, University of Göttingen, Nikolausberger Weg 23, D-37073 Göttingen, Germany

Abstract. It has been claimed (see Altmann 1976, Percus 1997) that focus particles cannot appear in cleft sentences if their meaning contradicts the generally assumed exhaustivity inference either via uttering it (e.g. *not only/also*) or presupposing it (e.g. *even*). The corpus data presented in this paper show, however, that it-clefts including non-exclusive focus particles do, in fact, appear in natural language examples for German. Additionally, this paper presents the results of a judgement experiment conducted on the basis of the natural language examples found in the corpus. The empirical data show that it-clefts which include non-exclusive focus particles are generally accepted by native speakers of German.

1 Introduction

A central question in the literature on it-clefts in English, as well as their counterparts in German and other languages, is whether the cleft structure (1) comes with the exhaustivity inference in (2), which is taken to be similar to the assertion in exclusive sentences (3).

(1) It was $[\text{Sue}]_F$ who climbed a mountain. (*Cleft Sentence*)

(2) Nobody other than Sue climbed a mountain. *(Exhaustivity Inference)*

(3) Only $[\text{Sue}]_F$ climbed a mountain.

While it is still an open question whether this exhaustivity effect is conventionally coded in the structure of the cleft and therefore semantic (see e.g. Büring and Križ 2013), or can be derived pragmatically as a conversational implicature (see e.g. Horn 1981), it is generally taken as a robust intuition that (2) can be derived from (1) (see e.g. Krifka 2008).

This paper will first take a general look into the semantic-pragmatic debate on the exhaustivity effect of it-clefts, and introduce Hungarian pre-verbal focus, a focus construction that has been analyzed in close relation to it-clefts. Previous studies which analyze corpus data for both it-clefts and Hungarian pre-verbal focus have argued in favor of a pragmatic approach to the exhaustivity claim, while theoretical approaches tend to suggest a semantic analysis. Within this debate, there are empirical findings regarding the special case of it-clefts which include focus particles, that are somewhat suprising: While semantic theories claim that certain focus particles can (or should) not appear in it-clefts, this paper presents corpus data that show that this is, in fact, possible. Taking the corpus data to indicate that focus particles are not at all bad in German it-clefts, this paper will then present a rating experiment that was designed in line with the natural language data from the corpus, and shows that certain it-clefts in combination with focus particles are judged as acceptable by native speakers.

2 Background

Despite a majority of the literature supporting the position that the inference in (2) can be derived from a cleft-structure, there is an ongoing debate on whether this exhaustivity inference is semantic (i.e. conventionally coded in the structure; cf. Percus 1997, Velleman et.al. 2012, Büring and Križ 2013, Halvorsen 1978) or pragmatic (i.e. a conversational implicature; see Horn 1981, 2014). There are several positions as to how the semantic exhaustivity is derived. Some theories hold that the effect is a conventional implicature (cf. Halvorsen 1978), while others analyze it as a uniqueness or maximality presupposition, parallel to definite descriptions (cf. Percus 1997), while still others take the effect as truth-functional, like the meaning of exclusive particles like *only* (cf. Atlas and Levinson 1981, É Kiss 1998), or as an exhaustiveness presupposition (cf. Büring and Križ 2013).

Empirical studies addressing the topic often support the pragmatic approach, as the exhaustivity effect appears to be cancellable (cf. Destruel 2012, Destruel et.al. 2015, Drenhaus et.al. 2011, DeVaugh-Geiss et.al. 2015). If the exhaustivity inference is semantic, it should not be (easily) cancellable. If clefts come with an exhaustivity inference which is pragmatic, however, this effect should be context-dependent and more easily cancellable.

2.1 Hungarian Pre-verbal Focus

From a cross-linguistical perspective, the exhaustivity effect in it-clefts is commonly analyzed on a par with Hungarian pre-verbal focus. If a focused expression appears in the immediately pre-verbal position in Hungarian, it is interpreted exhaustively, as in (4), in the same way as if it were in the scope of an exclusive like *only*. In cases where it appears in another position, as in (5), this exhaustiveness effect is not available (cf. Szabolcsi 1981, Onea and Beaver 2009).

(4) Péter MARIT csókolta meg.
Peter Mary.ACC kissed PRF
Peter kissed Mary (and no one else).

(5) Péter meg-csókolta MARIT.
Peter PRF-kissed Mary.ACC
Peter kissed Mary (and possibly someone else as well). (Onea and Beaver 2009: 342)

There is no comparable structural alternative to the pre-verbal focus in Hungarian. The strong exhaustivity of this focus construction has led to the tentative conclusion that the pre-verbal focus in Hungarian is semantically exhaustive, which makes it an ideal mode of comparison for the exhaustivity of it-clefts.

In a text-completion task for Hungarian focus, Onea and Beaver (2009) show that, while Hungarian pre-verbal focus does indeed come with an exhaustivity effect, it is not as strong as the effect which can be observed with the exclusive focus particle *csak/only*. These findings are further supported by the results of an extensive corpus study of Hungarian pre-verbal focus, in which Wedgwood et.al. (2006) show that this construction does not, in fact, show the robust exhaustivity inference that has been ascribed to it (e.g. by É Kiss 1998). Instead, the natural language examples from the Hungarian corpus used there show a great variety of focus adverbials with the focused element in their scope (*jórészt/ for the most part, legkevésbé/ least of all, elsőrban/ primarily, többek között/ among others*), which have the effect of explicitly de-exhaustifying the focused element.[1]

3 Focus Particles and it-Clefts

It has been claimed (cf. Altmann 1976, Percus 1997) that focus particles cannot appear in cleft sentences if their meaning contradicts the exhaustivity inference either via uttering it (e.g. *not only*) or presupposing it (e.g. *even*). However, the data presented in this paper show that it-clefts which include non-exclusive focus particles do, in fact, appear in natural language examples, as illustrated in (6) and (7):

(6) Es ist *auch* ihre Perspektivlosigkeit, die viele Jugendliche zur Flasche greifen lässt.
It is also their lack of perspective that makes many teenagers reach for the bottle.
(RHZ04/APR.20135 Rhein-Zeitung, 23.04.2004; Jugend braucht mehr Chancen)

(7) Es ist *vor allem* das Wetter, das uns bis jetzt einen Strich durch die Rechnung macht.
It is especially the weather that has messed up our plans so far.
(NUZ06/JUN.00081 Nürnberger Zeitung, 01.06.2006; Umsatzrückgang beim Einzelhandel im April - Wetter verregnete das Geschäft)

This study is focused on German it-clefts that include focus particles (e.g. exclusives like *only/nur*, additives like *too/auch*, iteratives like *again/wieder*, particularizers like *for example/beispielsweise*), of the kind illustrated in (6) and (7) above.

[1] While Wedgwood et.al. (2006) criticise a semantical analysis of the exhaustivity inference in Hungarian pre-verbal focus, it should be noted that they do not explicitly stress the relation between focus particles and the exhaustivity inference. This, in addition to the formal analysis of the different particles and implicatures, is left to further research.

Previous studies have shown the importance of considering corpus data in the context of the semantic-pragmatic debate on it-clefts and similar focus constructions (e.g. Wedgwood et.al. 2006 for Hungarian), as natural language examples seem to clash with primarily theoretical assumptions regarding the exhaustivity inference.

The corpus examples suggest that the focused element (or cleft pivot) is not the only item having the property denoted by the relative clause, hence (in some way) cancelling the exhaustivity inference, as illustrated in (8), where the conflict between exhaustivity inference (8a) and the meaning contribution of the particle (8b) can be clearly seen.

The newspaper article introduces Jette, a female Beagle, who did very well at a dog show, in which 18 dogs participated. The article then specifies why Jette did so well:

(8) Es ist *vor allem* Jettes Intelligenz, die verblüfft.
It is especially Jette's intelligence that is surprising.
(BRZ11/ MAI.01452 Braunschweiger Zeitung, 04.05.2011; Jette ist ein Superstar auf vier Pfoten)
a. *Expected exhaustivity inference of it-cleft*: Nothing besides Jette's intelligence is surprising.
b. *Presupposition of particle*: Something besides Jette's intelligence is (also) surprising.

Examples like this strongly suggest non-exhaustivity, which is incompatible with a semantic analysis of exhaustivity in it-clefts, and clash with the claim that clefts cannot include particles contradicting the exhaustivity inference (cf. Altmann 1976, Percus 1997).

Altmann (1976) claims that the function of a cleft is the emphasized identification combined with a uniqueness claim. Particles are only combinable with German it-clefts when they carry a scalar interpretation, and are not used to express non-exclusivity (like *nicht nur/ not only*), or presuppose non-exclusivity (like *sogar/ even*). The particle *nur/only* can therefore occur in cleft sentences, and *sogar/even* can occur in it-clefts in contexts which do not carry a non-exclusive reading. The particle *auch/also*, however, cannot occur in a cleft sentence because it carries a non-exclusive meaning in form of a non-uniqueness presupposition.

Furthermore, Percus (1997) argues that, since clefts of the form "It is [α] that φs" carry a requirement that $\forall x \varphi(x) \rightarrow x = \alpha$; a presupposition that only α has the property φ [2], they are incompatible with particles like *even* and *also*, and redundant (yet, possible, as in (11)) with exclusive particles such as *only.* Since these particles can usually associate with focus, as in (9), Percus (1997) argues that, in the case of it-clefts, the uniqueness-presupposition of the cleft and the semantics of the particles clash, resulting in unacceptable sentences, as illustrated below (10a–c):

[2] Percus notes that the presupposition is not as simple as this formula suggests. For further discussion, see Halvorsen 1978.

(9) a. It was even/also/only the case that [JOHN]$_F$ saw Mary. (Percus 1997: 341)

(10) a. ?It was even/also/only the case that it was [JOHN]$_F$ who saw Mary.
b. ??It was even [JOHN]$_F$ who saw Mary.
c. ??It was also [JOHN]$_F$ who saw Mary. (Percus 1997: 341)

Looking at (11), however, Percus (1997) notes that this structure with *only* does not lead to an unacceptability of the sentce, while the same construction and position of the particle lead to the mentioned unacceptability with *even* and *also*, as illustrated in (10b–c).

(11) It was only [JOHN]$_F$ who saw Mary.

Krifka (2008) observes that, if the exhaustivity inference in clefts is taken to be a presupposition as the effect of an identificational focus, an additive focus particle like *also* or *even* triggers a conflicting presupposition. Therefore, exhaustive focus cannot be taken to be compatible with additive particles, and he suggests a presuppositional or pragmatic analysis of exhaustivity instead of the common truth-functional one.

Horn (1981) presents examples which he takes to illustrate that the insertion of exclusives into a cleft does have a truth-functional effect, as illustrated in (12), while this is not the case for clefts without a more explicit way of stating the exhaustivity inference through the exclusive particle, as can be seen in (13).

(12) I know Mary ate a pizza, but I've just discovered that it was *only* a pizza that she ate.

(13) ??I know Mary ate a pizza, but I've just discovered that it was a pizza that she ate. (Horn 1981: 130)

Exclusives are taken to semantically assert exhaustivity, whereas the exhaustivity inference is taken to be a conversational implicature in the case of plain focus (see Beaver and Clark 2008).

Büring and Križ (2013) mention that in the case of (the possible, yet peculiar combination of) it-clefts with *only*, the exhaustivity presupposition of the cleft is tautologous. This resluts in it-clefts with *only* being equivalent to ordinary predication with *only*. They account for this by providing an analysis of it-cleft pivots of the form *only DP* as quantifiers.

Recent experimental studies support the claim that the exhaustivity effect in clefts is not truth-functional in the way that *only*-sentences are. Drenhaus et.al. (2011) provide data from questionnaire and on–line experiments which show that a violation of the exhaustivity effect in German *only*-sentences is less acceptable than in it-clefts. They conclude that the exhaustivity effect may have a different truth-functional status. This is supported by the results of a related ERP study, which showed different effects for exhaustivity violations in *only*-sentences and it-clefts, suggesting that the exhaustiveness violations in German it-clefts and *only*-sentences involve different processing mechanisms.

4 Corpus Evidence

A variety of examples for German it-clefts that include focus particles could be found in a non-exhaustive corpus study (COSMAS II search of several corpora including newspapers from German-speaking countries (Germany, Austria, and Switzerland) and Wikipedia entries, as well as Wikipedia forum discussions). The search was conducted by extracting examples of the form *"Es ist/war [x], der/die/das..." / "It is/was [x], that/which..."*. The random data collection amounts to nearly 400 German clefts which include focus particles. The examples with particles were found during a data collection for regular cleft structures, and were then collected and annotated separately to allow for further systematic research. Particular attention was paid to naturally occurring examples in which the exhaustivity inference is cancelled through the occurrence of a focus particle which has the clefted element in its scope, as well as to examples where the exhaustivity inference is strengthened through an exclusive.

The following focus particles were frequent in the corpus: *erneut, auch, beispielsweise, vor allem, nicht zuletzt, nur* (*again, also, for example, especially, not least, only*). Examples (14)-(20) illustrate the kinds of natural sentences that appear in the corpus. Sometimes, the German examples even include a combination of particles, as illustrated in (19) and (20).

(14) Es ist **nicht zuletzt** der strenge Rahmen aus Stein und Asphalt, der dem Central Park seinen einzigartigen Charakter verleiht.
It is ***not least*** *the rigid frame of rocks and asphalt that gives Central Park its unique character.*
(NZZ12/FEB.00616 Neue Zürcher Zeitung, 04.02.2012, S. 53; Geometrie der Gier Prisma der Welt)

(15) Es ist **vor allem** das Nahrungsangebot, das die Halden für Möwen attraktiv macht.
It is ***especially*** *the range of food that makes the dumps attractive for seagulls.*
(K00/MAI.35512 Kleine Zeitung, 04.05.2000, Ressort: Lokal; Möwenforscher lauert den Vögeln in Müllhalden auf)

(16) Es ist **nur** die in Udine gebotene Leistung, die momentan so nachdenklich stimmt.
It is ***only*** *the performance presented in Udine that makes one thoughtful right now.*
(A98/OKT.64464 St. Galler Tagblatt, 13.10.1998, Ressort: TB-SPO (Abk.); ≪Spiritus retour≫)

(17) Es ist **auch** der Reiz des Neuen, der viele hierher treibt.
It is ***also*** *the appeal of the new that brings many here.*
(BRZ06/MAI.10876 Braunschweiger Zeitung, 20.05.2006; Jobs im Ausland immer attraktiver)

(18) Es ist **zum Beispiel** der Ensemblespieler Alexander Seibt, der seine Karikatur eines menschlichen Aschenbechers in heftigst alkoholisiertem

Zustand zu einem der Höhepunkte des Abends macht.
*It is **for example** the ensemble member Alexander Seibt who turns his version of a heavily drunk human ashtray into a highlight of the evening.*
(NZS12/MAR.00217 NZZ am Sonntag, 11.03.2012, S. 63; Existenzialistische Flaschenpost)

(19) Es ist **zum Beispiel auch** Werner Langen, der die Sache im Europäischen Parlament als Berichterstatter massiv angeschoben hat.
*It is **for example also** Werner Langen that pushed the matter forward massively in the European Parliament as a messenger.*
(L98/DEZ.24042 Berliner Morgenpost, 05.12.1998, S. 6, Ressort: POLITIK; Über CDU pur und neue SPD)

(20) Es ist **vor allem auch** der politische Stil, der die Regierungsgegner erregt.
*It is **especially also** the political style that upsets the opposition.*
(K00/APR.32693 Kleine Zeitung, 22.04.2000, Ressort: Landespolitik; VP-Klubchef sorgt sich um "soziale Defizite" der Partei)

Contrary to the existing claims, the presented data show that there are, in fact, naturally occurring examples of it-clefts in combination with focus particles that have a non-exclusive meaning in German. A violation of the exhaustivity inference of the it-cleft (via non-exclusive focus particles like *auch/also*) does not support a semantic analysis of exhaustivity in it-clefts.

Returning to the example given in (8) above, recited here as (8'), another argument against a semantic anaylsis of it-cleft exhaustivity presents itself when looking at the sentence that follows the cleft.

> (8') Aus 18 Hunden stach die zweieinhalbjährige Beagledame hervor, und das nicht nur aufgrund ihres ungewöhnlichen Charmes oder wegen ihrer bezaubernden sherryfarbenen Augen.
> *Out of 18 dogs, the 2,5 year-old beagle lady was particularly noteworthy, and not just because of her exceptional charme or her lovely sherry-coloured eyes.*
> Es ist vor allem Jettes Intelligenz, die verblüfft.
> *It is especially Jette's intelligence that surprises.*
> Und ihre schnelle Auffassungsgabe.
> *And her fast understanding.*
> (BRZ11/ MAI.01452 Braunschweiger Zeitung, 04.05.2011; Jette ist ein Superstar auf vier Pfoten)

The focused element that is positioned in the cleft pivot (*Jettes Intelligenz*), is not intended to be the unique element which satisfies the property denoted by the relative clause. The following sentence states, on the contrary, another item that fulfills the cleft relative, thereby explicitly expressing non-exhaustivity.

5 Experimental Evidence

On the basis of this corpus data, a judgement experiment was conducted to gain systematic insights into the combinability of different focus particles and it-clefts in German. The aim was to collect data that allow for a comparison of different focus particles in different sentence environments. To achieve this, all particles were tested in the same sentence environments to determine whether some particles were more acceptable in clefts than others.

5.1 Method and Design

12 sentences were taken from the natural language examples found in the corpus and combined with 5 focus particles: *nur/only, auch/also, vor allem/especially, nicht zuletzt/not lastly, sogar/even*, which are all described as belonging to the same class of particles by Beaver and Clark (2008), who analyze these focus particles as conventionally associating with focus and bearing a lexically encoded dependency on focus. Each of the sentences was paired with each of the particles (including a condition with no particle) both in the clefted (21) and in the canonical version (22).

(21) Es ist die Einsamkeit, die die Menschen immer wieder an den Spieltisch treibt.
It is the loneliness that keeps bringing people to the gambling table.

(22) Die Einsamkeit treibt die Menschen immer wieder an den Spieltisch.
The loneliness keeps bringing people to the gambling table.

40 Participants (native German speakers, with an average age of 35 years) were asked to rate the sentences on a 7-point scale for acceptability (7 being fully acceptable, 1 being not acceptable). 4 participants were excluded from thre results as they did not complete the questionnaire. During a warm-up prior to the experiment phase, participants were presented with examples of poorly aceptable sentences which would be judged from 1 to 3, as well as examples of highly acceptable sentences which would be judged 5 to 7. A part of the filler items were designed to be rather unacceptable, in order to enable participants to make use of the whole range of the scale.

During the experiment, each sentence was only presented to each participant in one condition, ensuring that each participant only saw each sentence paired with one (or no) particle in order to avoid unwanted repetition effects. Partcipants saw a total of 24 sentences each, 50% critical items and 50% unrelated filler items. The experiment was conducted online using the free web-platform OnExp (https://onexp.textstrukturen.uni-goettingen.de/). The sentences were presented in written form individually on screen[3] and judged by checking a box with the matching number (1–7).

[3] Since the stimuli were presented to participants in written form, the present study did not control for the way participants interpreted the sentences, namely as focus-background or topic-comment clefts. This will be addressed in further research.

Below is a complete list of the clefted versions without focus particles that were taken from the corpus and presented to the participants in the different conditions. All these sentences were found in the corpus in combination with a focus particle and were then only marginally edited by replacing names with more recognizable ones.

(23) Es ist die Einsamkeit, die die Menschen immer wieder an den Spieltisch treibt.
It is the loneliness that keeps bringing people to the gambling table.

(24) Es ist Stephen Spielberg, der als Regisseur den Film vorantreibt.
It is Stephen Spielberg that presses the movie forward as the director.

(25) Es ist die Tiefe des Kraters, die die Wissenschaftler fasziniert.
It is the depth of the crater that fascinates the scientists.

(26) Es ist Michael Ballack, der es oft schafft, einen Ball in bedrängter Position anzunehmen.
It is Michael Ballack that often manages to receive a ball in a hard-pressed position.

(27) Es ist der Wiedererkennungswert, der die Ausstellung so reizvoll macht.
It is the recognition value that makes the exhibition so appealing.

(28) Es ist Seneca, der den zögernden Kaiser drängt, die Mutter zu beseitigen.
It is Seneca that urges the hesitant emperor to get rid of the mother.

(29) Es ist der medizinische Fortschritt, der die Kosten in die Höhe treibt.
It is the medical progress that increases the costs.

(30) Es ist Qaradawi, der auf Anfrage eines islamischen Armee-Seelsorgers mit anderen Gelehrten ein Fatwa für Muslime im amerikanischen Militär verfasste.
It is Qaradawi that issued a fatwa for muslims in the American army on demand from an islamic army counsellor together with other scholars.

(31) Es ist das Nahrungsangebot, das die Halden für Möwen attraktiv macht.
It is the range of food that makes the dumps attractive for seagulls.

(32) Es ist Jennifer Lawrence, die den Film zu einem Höhepunkt des Abends macht.
It is Jennifer Lawrence that turns the movie into a highlight of the evening.

(33) Es ist die Uniform der Feuerwehrleute, die die kleinen Kinder beeindruckt.
It is the firefighters' uniform that impresses the little kids.

(34) Es ist Werner Langen, der die Sache im Europäischen Parlament als Berichterstatter massiv angeschoben hat.
It is Werner Langen that pushed the matter forward massively in the European Partliament as a messenger.

5.2 Results and Discussion

The table and chart below present the average acceptability ratings for all conditions. In general, this experiment has shown that German it-cleft sentences are overall a little less acceptable than the canonical versions. This can possibly be explained by the fact that cleft sentences are not very frequent in (spoken) German and might seem less natural to speakers when presented to them in isolation.

Table 1. Average acceptability ratings of target sentences (all conditions)

	canonical	cleft
no particle	6,19	6,05
nur/only	5,62	5,86
auch/also	6,25	5,93
vor allem/especially	6,24	5,95
nicht zuletzt/not least	5,8	5,82
sogar/even	5,56	4,77

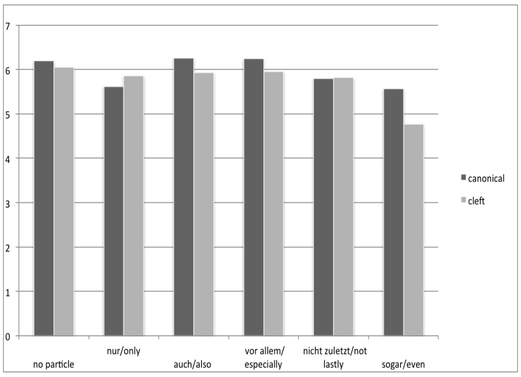


Fig. 1. Averange results per focus particle and sentence type (cleft/canonical)

However, this does not hold for the case of *nur/only*, where the rating is marginally better in the clefted condition. This might allow the conclusion that an exclusive focus particle in fact strenghtens the exhaustivity inference and therefore the general acceptability of the cleft sentence, thereby supporting Horn (1981) in his claim that the insertion of exclusives into a cleft has a truth-functional effect, while this is not the case for clefts without an exclusive particle which explicitly states exhaustivity.

In the case of *sogar/even* it can be said that the overall ratings were lowest in comparison to the other particles. For *sogar/even* and it-clefts, it might be the case that there is a mismatch between the speakers expectation to the context and the appearance of the focus particle in the cleft environment. Further research is needed to determine whether *sogar/even* is generally unacceptable to speakers when presented in an it-cleft. However, since the low ratings appear for both the clefted and canonical version in the case if *sogar/even*, it might be due to the fact that this particle is in general hard to process, as it is both an additive and a scalar particle (cf. König 1991).

Further statistical analysis is not necessary at this stage. The underlying hypothesis of this study, following the prior theoretical claims of Altmann (1976) and Percus (1997), was that it-clefts including non-exclusive focus particles would not be acceptable to German speakers, therefore leading to judgements between 1 and 3 on a 7-point scale. The total judgements are high enough to falsify this hypothesis without a more detailed statistical analysis at this point. Also, since participants judged those unrelated filler items that were expected to be rated between 1 and 3 (cf. examples (23)-(25) above) as unacceptable, it can be said that there is an observale effect, even without a statistical test for significance.

In general, the results show that German it-clefts are rated as acceptable by native speakers (above 5 on a 7-point scale). When combined with a non-exclusive focus particle, the acceptability ratings of the cleft sentences remain high and stable.

6 Conclusion

The approach of finding natural language evidence for non-exhaustive it-clefts is novel and not commonly used. Existing work using this method has shown, however, that it leads to important insights in the research on information structure and focus (e.g. Wedgwood et.al. 2006, Delin 1989). The data presented here build onto this foundation for the case of German.

In contrast to the claims of Altmann (1976) and Percus (1997), this study shows that German it-clefts can in fact occur in combination with a variety of (non-exclusive) focus particles, and native speakers judge them just as acceptable as cleft sentences without focus particles.

Additionally, German it-clefts combined with particles that carry a non-exclusive meaning *(auch/also, vor allem/especially, nicht zuletzt/not lastly)* were overall rated acceptable (above 5 on a 7-point scale). These findings stand in contradiction to the claim that it-clefts carry an exhaustivity inference which is conventionally coded into the structure, as this should not be overridden or cancelled by the insertion of a particle which takes scope over the focused element. Therefore, the presented data shed new light on the semantic-pragmatic debate regarding it-clefts, suggesting that (German) it-clefts are not semantically exhaustive.

References

Altmann, H.: Die Gradpartikeln im Deutschen. Tübingen: Max Niemeyer Verlag (1976)

Atlas, J.D. and S.C. Levinson: It-clefts, Informativeness and Logical Form: Radical Pragmatics. In: P. Cole (ed.): Radical Pragmatics. New York: Academic Press Inc, 1–61 (1981)

Beaver, D.I and B.Z. Clark: Sense and Sensitivity. How Focus determines Meaning. Oxford: Blackwell (2008)

Büring, D. and M. Križ: It's that and that's it! Exhaustivity and Homogeneity Presuppositions in Clefts (and Definites). Semantics and Pragmatics 6(6), 1–29 (2013)

Delin, J.: Cleft Constructions in English Discourse. University of Edinburgh (1989)

Destruel, E.: The French c'est-cleft: An empirical study on its meaning and use. In: Christopher Piñón (ed.): Empirical Issues in Syntax and Semantics 9: Selected Papers from CSSP, 95–112 (2012)

Destruel, E., D. Velleman, E. Onea, D. Bumford, J. Xue, and D. Beaver: A cross-linguistic study of the non-at-issueness of exhaustive inferences. In: F. Schwarz (ed.): Experimental Perspectives on Presuppositions. Springer, 135–156 (2015)

DeVeaugh-Geiss, J.P, M. Zimmermann, E. Onea and A.-C. Boell: Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. In: S. D'Antonio, M. Moroney, C. R. Little (ed.): Semantics And Linguistic Theory (SALT), Vol. 25, 373–393 (2015)

Drenhaus, H., M. Zimmermann, and S. Vasishth: Exhaustiveness effects in clefts are not truth-functional. Journal of Neurolinguistics 24, 320–337 (2011)

É. Kiss, K.: Identification Focus versus Information Focus. Language 74(2), 245–273 (1998)

König, E.: The Meaning of Focus Particles. A Comparative Perspective. London: Routledge (1991)

Halvorsen, P.-K.: Syntax and Semantics of Cleft Sentences. In: S. Mufwene, C. Walker, and S. Steever (ed.): Papers from the 12th Regional Meeting, Chicago Linguistic Society (1978)

Horn, L.: Exhaustiveness and the Semantics of Clefts. In: V. Burke and J. Pustejovsky (ed.): Papers from the 11th Annual Meeting of NELS, 124–142 (1981)

Horn, L.: Information Structure and the Landscape of (non-) at-issue Meaning. In: C. Fery and S. Ishihara (ed.): Handbook of Information Structure. Oxford: Oxford University Press (2014)

Kenesei, I.: On the Logic of Word Order in Hungarian. In: W. Abraham and S. de Mey (ed.): Topic, Focus and Configurationality. Amsterdam: J. Benjamins (1986)

Krifka, M.: Basic Notions of Information Structure. Acta Linguistica Hungarica 55, 243–276 (2008)

Onea, E. and D. Beaver: Hungarian Focus is not exhausted. In: E. Cormany, S. Ito, and D. Lutz (ed.): Proceedings of the 19th Semantics and Linguistic Theory Conference, 342–359 (2009)

Percus, O.: Prying open the Cleft. In: K. Kusomoto (ed.): Papers from the 27th Annual Meeting of NELS, 337–351 (1997)

Szabolcsi, A.: Compositionality in Focus. Acta Linguistica Societatis Linguistice Europaeae XV (1-2), 141–162 (1981)

Velleman, D., D. Beaver, E. Destruel, D. Bumford, E. Onea, and E. Coppock: It-clefts are IT (inquiry terminating) Constructions. In: A. Chereches (ed.): Proceedings of the 22nd Semantics and Linguistic Theory Conference (2012)

Wedgwood, D., G. Pethö, and R. Cann: Hungarian 'Focus Position' and English It-Clefts: The Semantic Underspecification of 'Focus' Readings (2006)

Ni-disjunction as a coordination marker and focus particle

Jovana Gajić

Georg-August-Universität Göttingen

This paper investigates the interpretation of Serbian *ni* and proposes a unified account for its role as a coordination marker, as well as a focus particle. *Ni* is analyzed as a strong NPI disjunction whose polarity sensitive behavior stems from obligatory exhaustification of alternatives which can only be successful in anti-additive environments, whereas the disjuncts are either overt or made up of the host proposition and a silent anaphor.

1 Coordination

In Serbian, the conjunction *i* is not restricted in its distribution with respect to the polarity of the sentence - it is grammatical both in a positive and in a negative sentence, as marked by the brackets on the verbal marker of sentential negation *ne* in (1). In contrast to this, *ni* is ungrammatical in a positive sentence (2). Both *i* and *ni* can appear either as single markers, thus introducing only the last member of the coordination, or they can be reiterated - one marker introducing each member of the coordination:

(1) Sofija (ne) piše (i) pesme i priče.
Sofija NOM (NEG) writes (and) poems ACC and stories ACC
'Sofija does(n't) write poems and stories'

(2) Sofija *(ne) piše (ni) pesme ni priče.
Sofija NOM NEG writes (ni) poems ACC ni stories ACC
'Sofija doesn't write poems or stories'

The interpretations that the above data get don't overlap entirely:

1. The sentence with *ni* (2) has only one reading (regardless of the number of *ni*s): 'Sofija doesn't write poems and she doesn't write stories'.
2. The negated version of the sentence with *i* (1) can have two readings:
   (a) 'Sofija doesn't write poems and she doesn't write stories'
   (b) 'Sofija doesn't write (both) poems and stories (only one of the two)'

Special coordination markers that emerge in negative contexts have not been extensively studied (de Swart 2001, Doetjes 2005, Wurmbrand 2008, Dagnac 2012, González and Demirdache 2015). The two central issues at stake are the question whether these items are inherently negative or semantically non-negative and the question whether they are best analyzed as conjunctions or as disjunctions. Focusing on two out of four logical combinations, a conjunction that introduces negative operators in each of its conjuncts and a disjunction in the scope of a negative operator, it is not trivial to tease one from the other option apart because one of the de Morgan equivalences (3) states that a conjunction outscoping negation is logically equivalent to a disjunction in the scope of negation.

(3) $(\neg p) \wedge (\neg q) = \neg(p \vee q)$

(4) a. 'Sofija doesn't write poems and she doesn't write stories' $(\neg p) \wedge (\neg q)$
b. 'Sofija doesn't write poems or stories' $\neg(p \vee q)$

(5) 'Sofija doesn't write (both) poems and stories' $\neg(p \wedge q)$

Ni in the example (2) can thus be interpreted as a conjunction (4a) or as a disjunction (4b), and the two readings are logically equivalent. The reading which is not possible with *ni* in (2) is the narrow scope conjunction one, as paraphrased in (5).

Arsenijević (2011) discusses these readings in some detail. In the same paper, he offers an analysis of Serbo-Croatian connectives, focusing on their morphological make-up and the syntax and semantics that can be derived from it. *Ni* is thus described as a negative conjunction (*n* - *i*(='and')).

1.1 Distribution

Ni can coordinate different kinds of constituents (DPs (6), NPs (8), PPs (9), VPs (10)). Single *ni* is bad with preverbal constituents.[1] Nevertheless, double *ni* is generally preferred in all positions in coordination.

(6) a. *(Ni) Sofija ni Lea ne ?ide/idu u školu.
ni SofijaNOM ni LeaNOM not goSg/goPl to schoolACC
'Neither Sofija nor Lea go to school'

b. Sofija nije upoznala ?(ni) mog brata ni tvoju
SofijaNOM didn't meetPART ni myACC brotherACC ni yourACC
sestru.
sisterACC
'Sofija didn't meet my brother or your sister'

(7) a. *(Ni) devojčice ni dečaci ne vole spanać.
ni girlsNOM ni boysNOM not likePl spinachACC
'Neither girls nor boys like spinach'

b. Sofija nije videla ???(ni) pse ni mačke.
SofijaNOM didn't seePART ni dogsACC ni catsACC
'Sofija didn't see (the) dogs or (the) cats'

[1] But even postverbal subjects coordinated by single *ni* yield strongly degraded sentences.

(8) a. ?* Ovaj (ni) prijatelj ni kolega nije dobar lingvista.
this$_{\text{NOM}}$ ni friend$_{\text{NOM}}$ ni colleague$_{\text{NOM}}$ isn't good$_{\text{NOM}}$ linguist$_{\text{NOM}}$
'This friend and colleague is not a good linguist'

b. Marko nije moj (ni) prijatelj ni kolega.
Marko$_{\text{NOM}}$ isn't my$_{\text{NOM}}$ ni friend$_{\text{NOM}}$ ni colleague$_{\text{NOM}}$
'Marko is neither my friend nor colleague'

(9) Sofija ne čuva knjige ?(ni) na polici ni u fijoci.
Sofija$_{\text{NOM}}$ not keep$_{\text{3Sg}}$ books$_{\text{ACC}}$ ni on shelf$_{\text{LOC}}$ ni in drawer$_{\text{LOC}}$
'Sofija doesn't keep books on the shelf or in the drawer'

(10) a. Lea nije (ni) pojela sendvič ni popila jogurt.
Lea$_{\text{NOM}}$ didn't ni eat$_{\text{PART}}$ sandwich$_{\text{ACC}}$ ni drink$_{\text{PART}}$ yogurt$_{\text{ACC}}$
'Lea didn't eat a/the sandwich or drink (the) yogurt'

b. Sofija neće (ni) sašiti ni kupiti haljinu.
Sofija$_{\text{NOM}}$ won't ni sew$_{\text{INF}}$ ni buy$_{\text{INF}}$ dress$_{\text{ACC}}$
'Sofija will neither sew nor buy a/the dress'

When it comes to bigger constituents, *ni* seems to be blocked from coordinating clausal structures, even in its multiplied *ni* incarnation[2]. Such structures become fully grammatical when gapped, as in (11c).

(11) a. ??? (Ni) Sofija nije videla Tamaru, ni Lea neće
ni Sofija$_{\text{NOM}}$ didn't see$_{\text{PART}}$ Tamara$_{\text{ACC}}$ ni Lea$_{\text{NOM}}$ won't
zvati Marka.
call$_{\text{INF}}$ Marka$_{\text{ACC}}$
'Sofija didn't see Tamara, nor will Lea call Marko'

b. ? (Ni) Sofija nije pojela sendvič, ni Lea nije
ni Sofija$_{\text{NOM}}$ didn't eat$_{\text{PART}}$ sandwich$_{\text{ACC}}$ ni Lea$_{\text{NOM}}$ didn't
popila jogurt.
drink$_{\text{PART}}$ yogurt$_{\text{ACC}}$
'Sofija didn't eat a/the sandwich, nor Lea drank (the) yogurt'

c. Lea nije videla Tamaru, ni Sofija Marka.
Lea$_{\text{NOM}}$ didn't see$_{\text{PART}}$ Tamara$_{\text{ACC}}$ ni Sofija$_{\text{NOM}}$ Marko$_{\text{ACC}}$
'Lea didn't see Tamara, nor Sofija (saw) Marko'

Importantly, regardless of the position of the *ni*-coordination in the sentence, all these examples contain sentential negation. This means that it is not possible to determine whether *(ni...)ni* is a disjunction in the scope of negation or a conjunction scoping over negation. For this task, more complex scopal configurations are needed.

[2] There is another coordination marker, *niti*, which is fully grammatical in such structures.

1.2 Disjunction or conjunction?

In order to tease apart the narrow scope disjunction from the wide scope conjunction interpretation of *(ni...)ni*, an additional scope-taking element can be added to the sentence. There are two possibilities: either inserting a quantificational expression that takes scope below the sentential negation (12), or inserting an expression that outscopes the negation (13):

(12) $\neg > \mathrm{Q} > \alpha \vee \beta$

(13) $(\mathrm{Q}\ \neg\alpha) \wedge (\mathrm{Q}\ \neg\beta)$

Availability of the interpretation represented in the configuration in (12) would provide evidence that *ni* can only be analyzed as a disjunction in the scope of negation, since it is not possible to transform this configuration into an equivalent one where *ni* would outscope negation (due to the scopal intervention of the third scope-taking element). Conversely, if the configuration in (13) is attested, this would mean that *ni* can only be analyzed as a conjunction that outscopes negative operators in each of its conjuncts, since there is no way to transform this LF into one where *ni* would be a narrow-scope disjunction.

Inspired by the so-called split-scope readings attested for Germanic negative indefinites (Penka 2010, Zeijstra 2011), a necessity modal that is outscoped by sentential negation makes it possible to check whether *(ni...)ni* is unambiguously a narrow scope disjunction (14). However, if the modal ends up outscoped by both the sentential negation and the *ni*-constituents, two equivalent interpretations are possible (15) again, so we are back to square one.

(14) $\neg > \Box\ [\alpha \vee \beta]$

(15) a. $\neg\ [[\Box\alpha] \vee [\Box\beta]]$ = b. $[\neg\Box\alpha] \wedge [\neg\Box\beta]$

For the Serbian example in (16), the availability of the reading in (14) would confirm that a narrow-scope disjunction analysis for *ni* is the correct one. At first glance, both interpretations, paraphrased below in (16b) and (16c) seem to be available.

(16) a. (Sofija) ne mora ni da kuva ni da čisti.
Sofija_{NOM} NEG has-to ni FIN cook ni FIN clean

b. (14): 'it is not necessary that Sofija cooks or cleans'

c. i. (15a): 'it is not the case that it is necessary for Sofija to cook or that it is necessary for Sofija to clean'

ii. (15b): 'it is not necessary that Sofija cooks and it is not necessary that Sofija cleans'

There is an entailment relation between the two scopal configurations: the one in (14) entails the ones in (15). This means that (16c) must be true whenever (16b) is true. It is thus necessary to verify if the only possible reading is (16b) actually, or (16c) is independently available (whether it can still be true when (16b) is false). A disambiguating scenario would be the one that is compatible with (16c), but incompatible with (16b). For example, Sofija's aunt owns a restaurant and she needs some extra work force, namely for cooking and cleaning, so Sofija's mother sends her over to help out during summer holidays. Thus, the mother obliged Sofija to work in aunt's restaurant on whichever of the two chores, which makes (16b) false in this context. At the same time, (16c) is true in this scenario because neither cooking nor cleaning was designated as a particular obligation to Sofija. The sentence in (16a) is not accepted by native speakers in this scenario, which means that the reading in (16c) can be dismissed. This provides evidence for a narrow-scope disjunction account of *(ni...)ni*, because the only available reading is the one (16b) where *(ni...)ni* cannot be represented as a wide scope conjunction.

Intervention with modals thus speaks in favor of analysing *(ni...)ni* as a disjunction in the scope of sentential negation. What needs to be further shown is that *ni* cannot be interpreted as a conjunction that introduces a negative operator in each of the conjuncts. For this purpose, a quantificational adverb will be used as a scope-intervening element, inspired by Shimoyama 2011. An adverb that outscopes sentential negation allows to test whether *(ni...)ni* is unambiguously a wide scope conjunction (17). When the adverb outscopes both the sentential negation and the *ni*-constituents, two equivalent interpretations are possible (18).

(17) $(\mathrm{Q}_{\mathrm{adv}}\ \neg\alpha) \wedge (\mathrm{Q}_{\mathrm{adv}}\ \neg\beta)$

(18) a. $\mathrm{Q}_{\mathrm{adv}} > (\neg\alpha \wedge \neg\beta) =$ b. $\mathrm{Q}_{\mathrm{adv}} > \neg(\alpha \vee \beta)$

If *(ni..)ni* is a conjunction that has negative operators in its scope, the interpretation in (17) should be available for (19a.

(19) a. (Sofija) obično nije (ni) kuvala ni čistila.
Sofija_{NOM} usually NEG.AUX3Sg ni cook_{PART} ni clean_{PART}

b. (17): 'It was usually not the case that Sofija cooked and it was usually not the case that Sofija cleaned'

c. i. (18a): 'It was usually the case that Sofija didn't cook and that Sofija didn't clean'

ii. (18b): 'It was usually not the case that Sofija cooked or cleaned'

Again, there is an entailment relation between the readings: the scopal configuration in (18) entails the one in (17). This means that, this time, what should be verified is whether the configuration represented in the paraphrase in (19b) is available independently form the other configuration (19c), i.e. in a context that is incompatible with the latter. Such a disambiguating scenario is given in the table in (20). This frequence of cooking and cleaning days would be compatible only with the interpretaion in (19b, since on four out of six days Sofija cooked and on four out of six days she cleaned. At the same time, (20) makes (19c) false because the distribution of the two activities is such that on only two out of six days Sofija did neither of the two. Crucially, the sentence in (19a) is not acceptable in the scenario depicted in (20).

(20)

	Mon	Tue	Wed	Thu	Fri	Sat
cooking	yes	no	no	no	no	yes
cleaning	no	no	yes	yes	no	no

This provides evidence against an analysis of *(ni..)ni* as a conjunction that outscopes sentential negation. Another argument against such an analysis comes from the observation that *(ni...)ni* is incompatible with collective predicates:

(21) * Ni Sofija ni Lea (ni Marko) se nisu sreli u
ni SofijaNOM ni LeaNOM ni MarkoNOM REFL didn't meetPART in
biblioteci.
libraryLOC

'Sofija, Lea and Marko didn't meet (each other) in the library.'

(22) * Ni Sofija ni Lea (ni Marko) nisu oformili tim.
ni SofijaNOM ni LeaNOM ni MarkoNOM didn't formPART teamACC

'Sofija, Lea and Marko didn't form a team (together).'

(23) Ni Sofija ni Lea (ni Marko) ne pišu projekte *zajedno.
ni SofijaNOM ni LeaNOM ni MarkoNOM not writePI projectsACC together

'Sofija, Lea and Marko don't write projects together'

Also, subjects coordinated by *ni* cannot be overtly modified with 'together' (23). This is unexpected for a conjunction marker, as they normally allow for non-Boolean interpretations with coordinated NPs/DPs. However, Sofija and Lea (and Marko) cannot be interpreted as a semantic plurality, as shown in the examples above.

Therefore, tests with necessity modals and quantificational adverbs, as well as incompatibility with collective readings, jointly show that *(ni...)ni* behaves as a disjunction in the scope of a negative operator and not as a conjunction that scopes over negative operators. Now, recall that sentences with *ni* were degraded when clausal coordination was involved, but fine with gapping (11]). Recent analyses of gapping (Coppock 2001, Johnson 2014) treat it as VP-ellipsis in a situation where what has been conjoined are the VPs beneath an auxiliary verb. This is consistent with an analysis of *ni* as a disjunction which needs to be in the scope of a negative operator, since in this case we would need only one negative operator that would scope over *ni*(s) and its disjuncts. Such distribution actually provides syntactic evidence for a disjunction-based account of *ni*, as it shows that *ni* must stay in the scope of a negative operator.

2 Strong NPI

The above-established fact that *(ni...)ni* always appears in the scope of negation makes it a good candidate for an NPI (Ladusaw 1992, Zeijstra 2004 inter alia).

However, NPIs are often grammatical in weaker, Downward Entailing (DE) environments[3]. *Ni*-coordination is ungrammatical in DE contexts, such as the scope of 'few', shown in (24). This means that *ni* cannot be analyzed as a weak NPI (Zwarts 1998).

(24) *Malo dece voli (ni) španać ni šargarepu.
few children$_{\text{GEN}}$ likes ni spinach$_{\text{ACC}}$ ni carrot$_{\text{ACC}}$
'Few children like spinach or carrots'

It is grammatical in anti-additive (25)[4] contexts and this makes it a suitable candidate for a strong NPI. It also means that *ni* is not a superstrong NPI, since such expressions are only grammatical in anti-morphic environments[5].

(25) Niko ne voli (ni) španać ni šargarepu.
ni-who$_{\text{NOM}}$ $_{\text{NEG}}$ likes ni spinach$_{\text{ACC}}$ ni carrot$_{\text{ACC}}$
'Nobody likes spinach or carrots'

Strong NPIs are often related to n-words in strict Negative Concord languages, where these items can co-occur with sentential negation in any number without yielding double-negation readings. In fact, a homophonous prefix (*ni-*) is combined with wh-expressions and this combination provides the class of n-words in Serbian. In addition, the analysis presented below predicts that there is only one disjunction operator at LF, thus, when *ni* is reiterated, those are just multiple realizations at PF and the position of *ni* doesn't necessarily coincide with its interpretation at LF. This could be implemented through a system of agreement à la Zeijlstra. Regardless of the exact label (n-word or strong NPI), the question is: what mechanism explains the distribution and the interpretation of *ni*?

2.1 Proposal

We take that Serbian *ni* is a semantically non-negative disjunction, whose polarity sensitive behavior stems from the presence of two formal features ([σ,D]) which need to be valued by matching features present on an operator $\text{O}^{\text{S}}{}_{[+\sigma,+\text{D}]}$ (Chierchia 2013). This silent operator c-commands the negative operator, as well as the *ni*-coordination and it performs exhaustification of alternatives similar in effect to focus particle 'only'. Once the agreement operation between $ni_{[-\sigma,-\text{D}]}$ and $\text{O}^{\text{S}}{}_{[+\sigma,+\text{D}]}$ is established and the valuing executed, the scalar (σ) and subdomain (D) alternatives are activated for the members of coordination introduced by *ni* and the operator $\text{O}^{\text{S}}{}_{[+\sigma,+\text{D}]}$ exhaustifies them. What this means is that all alternatives that are not entailed by the assertion have to be negated.

[3] These environments allow for inferences from sets to subsets: 'Few girls wore dresses' → 'Few girls wore blue dresses'.

[4] These environments satisfy the equivalence: $f(X \cup Y) \Leftrightarrow f(X) \cap f(Y)$; for example - 'No girls sang or danced' is equivalent to 'No girls sand and no girls danced'.

[5] These environments satisfy both the equivalence $f(X \cup Y) \Leftrightarrow f(X) \cap f(Y)$ and the equivalence $f(X \cap Y) \Leftrightarrow f(X) \cup f(Y)$; for example 'Girls didn't sing or dance' is equivalent to 'Girls didn't sing and girls didn't dance' and 'Girls didn't sing and dance' is equivalent to 'Girls didn't sing or girls didn't dance'.

(26) For the example in (2):

a. Assertion: $O^S\neg(\alpha\vee\beta)$

b. Scalar (σ) alternatives: $\neg(\alpha\wedge\beta)$

c. Subdomain (D) alternatives: $\neg\alpha$, $\neg\beta$

d. After EXH: $\neg(\alpha\vee\beta)$

e. where α = 'Sofija writes poems' and β = 'Sofija writes stories'

As shown in (26), exhaustification is vacuous for any sentence with *ni* where sentential negation, marked by the presence of a verbal marker (*ne/ni*-AUX), is correctly realized. This is because the assertion ('it is not the case that Sofija writes poems or that Sofija writes stories', as in (26a)) entails all the alternatives, be it scalar ('it is not the case that Sofija writes poems and that Sofija writes stories', as in (26b)) or subdomain ('it is not the case that Sofija writes poems', 'it is not the case that Sofija writes stories', as in26c). Crucially, in a negative sentence exhaustification does not yield a contradiction, unlike in its positive counterpart, shown in (27):

(27) For the example in (2):

a. Assertion: $O^S(\alpha\vee\beta)$

b. Scalar (σ) alternatives: $\alpha\wedge\beta$

c. Subdomain (D) alternatives: α, β

d. After EXH: $(\alpha\vee\beta) \wedge \neg(\alpha\wedge\beta) \wedge \neg\alpha \wedge \neg\beta$

e. where α = 'Sofija writes poems' and β = 'Sofija writes stories'

When *(ni...)ni* is embedded in a positive environment, the assertion (27a) is the weakest of all the alternatives, so scalar (27b) and subdomain (27c) alternatives have to be negated. This exhaustification of alternatives (27d) yields a contradiction with respect to the assertion and renders the sentence ungrammatical.

Sets of scalar and subdomain alternatives, whose activation turned out to be crucial in the account for restrictions in the distribution of *ni*, are posited for the plain disjunction as well (Sauerland 2004, Fox 2007). The major difference is that, in the case of Polarity Sensitive Items such as *ni*, these alternatives must always be active and exhaustified. This is why *ni* is only acceptable in negative contexts, whereas 'or', for example, does not have a restricted distribution.

The very restricted distribution of *ni* (AA contexts only) is explained by invoking strong exhaustification of alternatives (Gajewski 2011). Performed by O^S, this strong exhaustification takes into account not only the truth-conditional content of the sentence that it embeds, but also the presuppositions and the implicatures that arise. This accounts for the descriptive generalization that *(ni...)ni* is ungrammatical in DE contexts which are not AA, such as the scope of 'few' in (24) - 'few children like x' bears an implicature that 'some children like x' and this disrupts the downward monotonicity of the sentence, yielding the exhaustification contradictory.

3 Additive focus particle

The above-presented disjunction-based analysis of Serbian *ni* can be extended to its other roles in grammar. In the previous section it was shown that *ni* can introduce overt constituents, acting as a coordination marker. Nevertheless, *ni* can also serve as an additive focus particle restricted to negative contexts, analogous to English 'either'. In this incarnation, *ni* attaches to a focalized constituent (the subject 'Lea' in (28), the VP 'do homework' in (29), or the PP 'on the shelf' in (30)) and the corresponding set of alternatives ((28a) for (28b), (29a) for (29b), (30a) for (30b)) can be activated.

(28) a. 'Sofija didn't do the homework', 'Marko didn't do the homework', etc.

b. Ni Lea nije uradila domaći.
ni LeaNOM didn't doPART homeworkACC
'Lea didn't do the homework, either'

(29) a. 'She didn't wash the dishes', 'She didn't feed the dog', etc.

b. Nije ni uradila domaći.
didn't ni doPART.F homeworkACC
'She didn't do the homework, either'

(30) a. 'It's not in the drawer', 'It's not in the bag', etc.

b. Nije ni na polici.
isn't ni on shelfLOC
'It's not on the shelf, either'

The data above exemplify syntactically grammatical occurrences of the additive focus particle *ni*. However, there is an anaphoric requirement born by *ni*, due to which these examples would be infelicitous without corresponding contextual alternatives that can serve as an antecedent. Extending the disjunction-based analysis proposed for coordination facts, and in line with Ahn's (2014) account of English focus particle 'either', Serbian *ni* is understood as a disjunction that takes as its arguments the host proposition and a silent anaphor. It is the silent anaphor that requires a negative antecedent (Rullmann 2003). This antecedent must be distinct from the host proposition (Kripke 2009) and at least one of the alternatives from the set in the focus value of the host must be entailed by it. The distributional restrictions to which *ni* is subject even as an additive focus particle are predicted by its disjunctive nature.

(31) For the example in (28b)

a. Assertion: $\mathrm{O}^{S}\neg(\mathrm{p}\vee\mathrm{q})$

b. Scalar (σ) alternatives: $\neg(\mathrm{p}\wedge\mathrm{q})$

c. Subdomain (D) alternatives: $\neg\mathrm{p}$, $\neg\mathrm{q}$

d. After EXH: $\neg(\textit{ni}\ \mathrm{p}) = \neg(\mathrm{p}\vee\mathrm{q})$

e. where p = 'Lea did the homework' and $q\in[[\mathrm{p}]]^{F}$

Again, the polarity sensitivity of the disjunction *ni* is lexicalized through formal features ([-σ] and [-D]) which are responsible for activating the sets of scalar and domain alternatives (respectively). Once these features are valued by covert O^S, the exhaustification of alternatives is performed and no contradiction arises because a negative environment makes the assertion stronger than any other alternative, scalar or subdomain. The covert ONLY-operator and sentential negation are thus the key ingredients not only when *ni* marks a coordination with overt disjuncts, but also when one of the two disjuncts is a silent anaphor. This highlights the prerequisite for *ni* to be a disjunction in the scope of a negative operator - this way both disjuncts, i.e. alternatives representing the normal semantic value (the host proposition) and the focus value (vie the silent anaphor), get negated. Furthermore, additive focus particle *ni* is unacceptable in a sentence without negation for the same reason for which *ni*-coordination is - as shown in the previous section, a contradiction arises between the assertion and exhaustified alternatives because the ONLY-operator must negate all alternatives that are not entailed by the assertion. In a positive (non-AA) sentence, scalar and subdomain alternatives are stronger than the assertion, however, their exhaustification is in direct collision with the (positive) assertion.

4 Scalar focus particle

When used as a focus particle, with one silent disjunct, *ni* can also express a scalar meaning, different from the one described for additive *ni* ('either') and similar to that of English 'even' in a negative sentence.

(32) a. 'Lea did the homework' $>_\mu$ 'Sofija did the homework' $>_\mu$ 'Marko did the homework' $>_\mu$,etc.

b. Ni Lea nije uradila domaći.
ni Lea$_{\text{NOM}}$ $_{\text{NEG.AUX3Sg}}$ do$_{\text{PART}}$ homework$_{\text{ACC}}$
'Not even Lea did the homework'

c. i. 'She didn't wash the dishes', 'She didn't feed the dog', etc.

ii. Nije ni uradila domaći.
didn't ni do$_{\text{PART.F}}$ homework$_{\text{ACC}}$
'She didn't do the homework, either'

d. i. 'It's not in the drawer', 'It's not in the bag', etc.

ii. Nije ni na polici.
isn't ni on shelf$_{\text{LOC}}$
'It's not on the shelf, either'

This change in the meaning for the same examples ((28b)=(32b), (29b)=(32c-ii), (30b)=(32d-ii)) can be captured by switching from the ONLY type of exhaustification to the EVEN type (Chierchia 2013). The motivation for this switch is that active focus alternatives are no longer sufficient by themselves to make the sentence felicitous, these alternatives also need to be ordered on a likelihood scale. This is expressed with the probability measure μ in (33c).

(33) For the example in (32b):

a. Assertion: $\neg$(p$\vee$q)

b. Scalar (σ) and subdomain (D) alternatives: p $<_{\mu}$ q

c. After EXH: $\neg$((p$\vee$q) $\vee$ p$<_{\mu}$q) = $\neg$p $\wedge$ $\neg$q $\wedge$ $\neg$p$<_{\mu}\neg$q

d. where p = 'Lea did the homework' and $q \in [[\mathrm{p}]]^F$

The O$^{\mathrm{S}}$ operator can exhaustify the alternatives introduced by *ni* without yielding a contradiction, in a negative sentence. However, the covert ONLY-operator cannot perform the exhaustification of such linearly ordered alternatives and render the additional component of meaning in (32b) - Lea not doing her homework was the least likely alternative. On the other hand, EVEN-exhaustification can capture the emphatic effect (absent with O-exhaustification), thanks to the introduction of the probability measure. The reason why alternatives now look slightly different is that, as shown in (33b), what used to be subdomain alternatives, i.e. individual overt disjuncts (p, q), are now the host proposition (p) and the silent anaphor (q), ordered on a likelihood scale. As a polarity sensitive disjunction that bears the [σ,D] feature, *ni* gets checked and valued by the E$^{\mathrm{S}}$ operator and this activates parallel and logically stronger alternatives, ordered with respect to some contextually relevant probability measure ($<_{\mu}$).

The E-operator is thus invoked to signal that the assertion is the least likely among the relevant alternatives. This correctly captures the meaning of the scalar focus particle *ni* in a negative context (32b) - it is not the case that Lea did the homework, and it is not the case that someone else did the homework, and it is not the case that Lea doing the homework was the least likely alternative. We get the 'not even' meaning for the scalar *ni.*

5 Conclusions

In our proposal, Serbian *ni* is a strong NPI disjunction always in the scope of sentential negation and this is valid for both its coordination and focus particle incarnations. This correctly predicts the polarity sensitive behavior of *ni*, via the alternatives and exhaustification framework. Outside of the scope of the present paper is a possible parallel with the plain, distributionally non-restricted, conjuction *i*, which also exhibits focus particle behavior in Serbian, and thus represents a natural extension of this research.

References

Ahn, D. (2014): The semantics of additive *either.* Proceedings of SuB 19, eds. Csipak, Zeijlstra.

Arsenijevic, B. (2011): Serbo-Croatian coordinative conjunctions at the syntax-semantics interface. The Linguistic Review 28, 175–206.

Chierchia, G. (2013): Logic in Grammar; Polarity, Free Choice and Intervention. Oxford Studies in Semantics and Pragmatics 2, OUP.

Coppock, E. (2001): Gapping: in defense of deletion. Proceedings of the CLS 37, 133–147.

Dagnac, A. (2012): Gapping as vP-coordination: an argument form French strict NPI licensing. Ellipsis, University of Vigo.

deSwart, H. (2001): Négation et coordination: la conjonction *ni*. Adverbial Modification, eds. Bok-Bennema, de Jonge, Kampers-Manhe, Molendijk.

Doetjes, J. (2005): The chameleonic nature of French *ni*: negative coordination in a negative concord language. Proceedings of SuB 9, eds. Maier, Barry, Huitnik.

Gajewski, J. (2011): Licensing strong NPIs. Natural Language Semantics 19(2), 109–148.

Gonzalez, A., Demirdache, H. (2015): Negative coordination: single vs. recursive *ni* in French. Proceedings of 44th LSRL, Western University.

Johnson, K. (2014): Gapping. MS, UMass Amherst.

Kripke, S.A. (2009): Presupposition and anaphora: Remarks on the formulation of the projection problem. Linguistic Inquiry 40(3), 367–386.

Rullmann, H. (2003): Additive particles and polarity. Journal of Semantics 20(4), 329–401.

Sauerland, U. (2004): Scalar implicatures in complex sentences. Linguistics and Philosophy 27(3), 367–391.

Shimoyama, J. (2011): Japanese Indeterminate Negative Polarity Items and their scope. Journal of Semantics 28, 413–450.

Wurmbrand, S. (2008): *Nor: Neither* disjunction *nor* paradox. Linguistic Inquiry, 511–522.

Zeijstra, H. (2011): On the syntactically complex status of negative indefinites. Journal of Comparative German Linguistics, 111–138.

Penka, D. (2010): Negative Indefinites. Oxford University Press.

Ladusaw, W. (1992): Expressing Negation. SALT 2 Proceedings, Columbus: The Ohio State University.

Zeijstra, H. (2004): Sentential Negation and Negative Concord. LOT. University of Amsterdam.

Zwarts, F. (1998): Three types of polarity. Kluwer Academic Publishers, The Netherlands, 177–238.

Fox, D. (2007): Free choice disjunction and the theory of scalar implicatures. Presupposition and Implicature in Compositional Semantics, eds. Sauerland U. and P.Stateva.

Exhaustivity in Mandarin *Shi . . . (de)* Sentences: Experimental Evidence

Ying Liu[1] and Yu'an Yang[2]

[1] City University of Hong Kong
[2] Chinese University of Hong Kong

Abstract. In this study, we present three experiments evaluating different hypotheses regarding the exhaustivity of Mandarin *shi . . . (de)* cleft construction (SD). Experiment 1 shows that the exhaustivity of this construction is received differently from restrictive particle *zhiyou* and plain focus sentences; Experiment 2 further demonstrates that exhaustivity of SD cannot be canceled by contradiction continuation led by *In fact . . .*; finally, Experiment 3 indicates that the existential meaning of SD can project over negation while exhaustivity cannot. These results suggest that the exhaustivity of SD may not be directly asserted, conversationally implied, nor presupposed. Thus, we are directed back to an epiphenomenal proposal in line with [19] and [11].

1 Introduction

Shi . . . (de) construction (henceforth SD) has long been recognised as the Mandarin counterpart of English *it*-cleft, as illustrated in (1) [3] ([20] among others). This construction in both English and Chinese encodes three meaning components, i.e. existential meaning, identificational meaning and exhaustivity.

(1) Shi [Xiaogao he Xiaopang]$_F$ chidao le.
SHI Xiaogao and Xiaopang late ASP
'It is Xiaogao and Xiaopang who were late.'
Existential presupposition: There is someone who was late.
Identificational assertion: Xiaogao and Xiaopang were late.
Exhaustivity: Besides Xiaogao and Xiaopang, no one else was late.

It is generally agreed that cleft constructions in English and many other languages place the first two meaning components in presupposition and assertion respectively, but exhaustivity triggered much debate. The semantic account places the exhaustivity in assertion (e.g. [7]) or in presupposition (e.g. [19], [3]), while the pragmatic account takes it as a conventional implicature ([9]) or as a conversational implicature (e.g. [13], [6]). Although many scholars have analysed the Mandarin SD cleft sentence and agreed that it also has existential

[3] Glosses: ASP: aspectual marker, LOC: localizer, CL: classifier, SHI...(DE): the cleft construction in Mandarin

presupposition and identificational assertion (e.g. [16]), the status of exhaustivity, especially on which layer of meaning SD encodes exhaustivity has rarely been discussed.

This study sets out to investigate clefts' exhaustivity with experimental data in Mandarin. Specifically, we wish to address two issues: (i) what is the status of exhaustivity in SD; is it encoded in presupposition, assertion, or implicature? (ii) how is exhaustivity derived in SD, and why?

In what follows, we will first review the semantic and pragmatic accounts of clefts' exhaustivity that motivate this current experimental investigation. Section 3 to 5 present three experiments targeting the assertion, conversational implicature and presupposition analysis of SD's exhaustivity. Finally, we discussed a possible analysis to our experimental results.

2 Background

Besides clefts, restrictive particles like *only* as in (2) and plain focus sentences (henceforth PF) like B's answer to a *wh*-question in (3), also infer exhaustivity.

(2) Only [Mary]$_F$ was late.

(3) A: (Among Mary, Peter, and Susan,) who was late?
B: [Mary]$_F$ was late.

Previously, scholars have reach a consensus that the exhaustivity of a restrictive particle like *only* is asserted ([7] among others) while that of plain focus sentences is conversationally implicated (e.g. [18]). For example, in (2) the sentence asserts that besides Mary, nobody else was late, but in (3) the same meaning is implied. Cleft sentences, on the other hand, received much controversies, which we will take a closer look now.

2.1 Assertion Analysis of Clefts' Exhaustivity

Based on the similarities between clefts and exclusive *only*, É. Kiss ([7]) among others propose that the exhaustivity of clefts is part of its assertion. Lee ([16]) applies this analysis to Chinese *shi . . . (de)* clefts:

(4) Shi [Zhangsan]$_F$ da Lisi de.
SHI Zhangsan beat Lisi DE
"It was Zhangsan that beat Lisi."
Presupposition: 'Someone beat Lisi.'
Assertion: The 'someone' equals Zhangsan; Except Zhangsan, there are no other people who beat Lisi.' ([16, p.95])

2.2 Conversational Implicature Analysis of Clefts' Exhaustivity

Observing the disparity between the exhaustivity of *it*-clefts and that of *only*, Horn ([13]) proposes that clefts' exhaustivity is a generalized conversational implicature, calculated from the Maxim of Quantity. This proposal found support in recent experimental studies (e.g. [4], [5], [6]). These studies show that (i) under certain contexts, cleft sentences accept non-exhaustive interpretation ([4], [18] among others); (ii) contradicting clefts' exhaustivity is processed differently from contradicting the assertion or presupposition content of *only* ([6]).

2.3 Presuppositional and Conventional Implicature Analysis of Clefts' Exhaustivity

Drawing on the close relationship between definiteness and exhaustivity, Percus ([19]) and Hedberg ([10], [11]) propose that the exhaustivity of clefts is derived from the maximality of a definite DP formed by the cleft pronoun *it* and the cleft clause. Büring ([2]) argues that exhaustivity is realised as a conditional (e.g. "if Xiaogao and Xiaopang were late, no one else was late" for (1)) in the presupposition of clefts. Since the assertion ("Xiaogao and Xiaopang were late") made the antecedent of this conditional true, exhaustivity ("no one else was late") is thusly derived. Later, Büring and Križ ([3]) observed that *it is x that P* should presuppose "x is not a proper part of the maximal member of P" ([3, p.4]), and revised this conditional into a homogeneity presupposition. Under this account, (1) presupposes that the plural entity [Xiaogao and Xiaopang] is not a proper part of the sum of all the individuals being late; i.e. either Xiaogao and Xiaopang were the only individuals being late or they were not late at all. Combined with the assertion, the second conjunct was falsified, so Xiaogao and Xiaopang were the only people who were late and the exhaustivity of (1) is derived. Velleman et al. ([21]) propose that both clefts and *only* are inquiry terminating constructions that have two focus sensitive operators MAX and MIN: the former specifies that "no true answer is strictly stronger than p" while the latter states that "There is a true answer at least as strong as p." While *only* presupposes MIN and asserts MAX, clefts assert MIN and presuppose MAX.

Halvorsen ([9]) argues for a conventional implicature analysis of the exhaustivity of *it*-clefts. According to her, both the exhaustiveness implicature and existential implicature are computed on the basis of an intermediate structure which is unaffected by negation on copula. Therefore, both meaning components could survive in negated and questioned clefts.

In summary, the assertion proposal draws an analogy between *zhiyou* and SD, which predicts that exhaustivity affects the truth-condition of these two structures in the same way. As for the conversational analysis hypothesis, it would predict that the exhaustivity of SD is comparable to that of PF regarding the diagnostics of conversational implicatures. If the exhaustivity of SD is presupposed or conventionally implicated, it should survive projective contexts like negation, conditional antecedent and modals ([14], [15]).

3 Experiment 1

Experiment 1 compared Mandarin speakers' acceptance to exhaustive inference in *shi . . . (de)* clefts (SD) with sentences containing *zhiyou* (ZY) and plain focus sentences (PF). As discussed above, ZY asserts while PF conversationally implicates exhaustivity. In a neutral context, speakers should assign a higher degree of acceptance to asserted exhaustivity than to exhaustivity encoded in other layers of meaning, while conversationally implied exhaustivity may not even arise and thus should receive a relatively low score. Using ZY and PF as reference, we could have a peek into the nature of clefts' exhaustivity.

Methods This experiments employed a inference judgment task presented as a web-based questionnaire. Sixty-one speakers of Mandarin Chinese (age: 23 to 58, mean 31) were first introduced to David, a fictional non-native speaker of Mandarin, and then asked to judge on a scale from 1 to 5 (1 being the least acceptable) how acceptable is David's inference in a given scenario. Each scenario consisted of a short background as lead-in, a pre-recorded statement made by a Mandarin-speaking friend of David's as the eliciting sentence, and finally David's inference of this statement as target inference.

Twelve sets of scenarios were created. Each eliciting sentence underwent four permutations: *zhiyou* "only" sentences (ZY), *shi ... (de)* cleft sentences (SD), plain focus sentences (PF), and simple SVO sentences without any focus (referred to as canonical sentence, CN). An example is given in Tab. 1. Together the 96 items were assigned to six lists in a Latin square fashion. The sixteen items in each list was pseudo-randomized with thirty-six filler items. All audio stimuli and inference sentences were verified as grammatical by two native Mandarin speakers, so participants' judgment would not be interfered by grammaticality.

Dian li, hongcha maiwan le. 'In the store, black tea was sold out.'	CN
Dian li, [hongcha]$_F$ maiwan le. 'In the store, [black tea]$_F$ was sold out.'	PF
Dian li, **shi** [hongcha]$_F$ maiwan le. 'In the store, it is [black tea]$_F$ that was sold out.'	SD
Dian li, **zhiyou** [hongcha]$_F$ maiwan le. 'In the store, only [black tea]$_F$ was sold out.'	ZY
David's Inference: So other drinks were not sold not.	

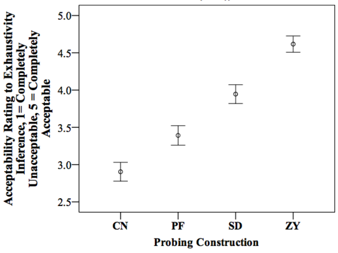
Table 1. Four permutations of a eliciting sentence and the target inference in Experiment 1

Predictions (i) Following the assertion analysis, the acceptability of SD's exhaustivity should pattern with that of ZY; (ii) following the conversational implicature analysis, the acceptability of SD's exhaustivity should pattern with that of PF; (iii) if the exhaustivity of SD is encoded otherwise, its acceptability should pattern with neither constructions.

Results Results from sixty complete questionnaires were analyzed. The mean acceptability ratings of exhaustive inference of the four types of probing constructions are presented in Fig. 1. One-way ANOVA reveals that the difference among the four probing constructions is statistically significant (F=137.9,

p=0.000). A post-hoc Bonferroni test suggests that the mean acceptability to exhaustive inference of SD (mean=3.95) was significantly lower than that of ZY (mean=4.62, p=0.000), while higher than that of PF (mean=3.39, p=0.000). These three constructions all received a higher acceptability to exhaustivity than CN (mean=2.90, p=0.000).

Fig. 1. Exhaustivity in four types of sentences (means with confidence intervals 95%)



Discussion This experiment helps to paint a general picture of how well exhaustivity inference of various exhaustivity-inducing constructions is received among Mandarin speakers. While PF, SD, and ZY sentences all elicit an exhaustive interpretation, the levels of acceptance vary, suggesting that the status of exhaustivity of the three tested types of sentences differs from each other. Results from our experiment then fail to support the assertion and conversational implicature analysis of SD's exhaustivity, as SD patterned with neither ZY nor PF regarding the acceptability of exhaustive inference.

4 Experiment 2

One of the hallmarks of a conversational implicature is its cancelability, i.e. it may be suspended under certain contexts ([8]). As illustrated in (5), PF (5b) but not SD (5a) is compatible with a non-exhaustive context introduced by *biru*, "for example", suggesting that the exhaustivity of PF is suspended in this context. SD's exhaustivity, on the other hand, cannot be suspended, indicating that it may not be a conversational implicature.

(5) *Context: The tutor finished grading last week's quiz. The lecturer asked:*
Lecturers: Who didn't pass the exam?
TA: Many students didn't pass,

a. ?? biru, shi [Zhangsan]$_F$ bu jige.
For example, SHI Zhangsan not pass
'For example, it is Zhangsan who didn't pass.'

b. biru, [Zhangsan]$_F$ bu jige.
For example, Zhangsan not pass
'For example, Zhangsan didn't pass.'

This pattern motivates the current experiment to test the conversational implicature analysis of SD's exhaustivity. The other diagnostic for cancelability is that a conversational implicature can be canceled by its following utterance ([8]; see [17] for a recent discussion). For example, if an utterance conversationally implicates p, a follow-up like *In fact (not p)* can override p. In this experiment, we added such a follow-up to SD. If the exhaustivity of SD can be canceled, an SD utterance with the follow-up *In fact, someone else did it too* would still be acceptable. Same as Experiment 1, ZY and PF were set as reference.

Method This experiment adopted a felicity judgment task presented as a web-based questionnaire. Thirty-six Mandarin speakers (age: 21-36, mean: 25.7) were asked to judge whether David's utterance in each scenario was acceptable on a scale from 1-5 (same as Experiment 1). Each scenario consisted of a short background, a question by David's Mandarin-speaking friend as elicitation, and finally a pre-recorded David's response to the question as the target sentence.

Nine sets of testing scenarios were created, each with three permutations on the target sentence: ZY, SD, and PF. Each target sentence was composed of two conjuncts: the first varied with constructions, and the other was the follow-up *In fact, someone else did it too.* The elicitation question also contained two parts: the first identified all the alternatives with a prepositional phrase to create an exhaustivity-inducing context, followed by a *wh*-question. An example is given in Tab. 2. All items were verified by two native speakers; specifically we wanted to make sure that the first conjunct of each target sentence was an felicitous answer to the eliciting question. The testing and filler scenarios were then assigned to three lists in a Latin square fashion, such that each list displayed nine testing scenarios and nine filler scenarios in a pseudo-randomized order.

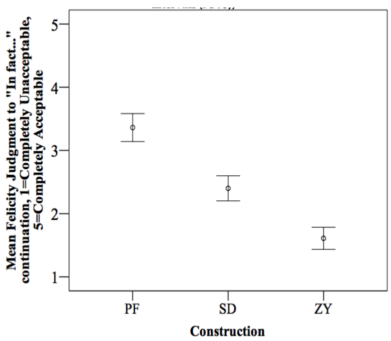
***Wh*-question: Between Mo Yan and Yu Hua, who has won the prize?**	
[Mo Yan]$_F$ na guo jiang; shishishang, Yu Hua ye na guo jiang. [Mo Yan]$_F$ has won the prize; in fact, Yu Hua also has won the prize.	PF
Shi [Mo Yan]$_F$ na guo jiang; shishishang, Yu Hua ye na guo jiang., It is [Mo Yan]$_F$ who has won the prize; in fact, Yu Hua also has won the prize.	SD
Zhiyou [Mo Yan]$_F$ na guo jiang; shishishang, Yu Hua ye na guo jiang. Only [Mo Yan]$_F$ has won the prize; in fact, Yu Hua also has won the prize.	ZY

Table 2. Example of three permutations of target sentence from Experiment 2

Predictions (i) Following the conversational implicature analysis, the cancelability of SD's exhaustivity should pattern with that of PF; (ii) if the exhaustivity of SD is encoded otherwise (e.g. in assertion or in presupposition), its cancelability should not pattern with PF.

Results Thirty-five complete questionnaires were included in the analysis. The mean acceptability ratings of the three types of sentences are presented in Fig. 2. There was a statistically significant difference among constructions as determined by one-way ANOVA (F=76.345, $p < 0.01$); a Bonferroni test revealed that the acceptability to cancelation continuation of PF (mean = 3.4) was significantly higher than that of SD (mean = 2.4, $p < 0.0001$) and ZY (mean = 1.6, $p < 0.0001$); SD and ZY also differ ($p < 0.0001$).

Fig. 2. Cancelling exhaustivity in three structures (means with confidence intervals 95%)



Discussion This experiment shows that SD differs from ZY and PF regarding the cancelability of exhaustivity. Horn ([12]) as well as Zimmermann and Onea ([22]) have argued that the difference between the exhaustivity of clefts and in situ prosodic PF is connected to the existential presupposition. However, in our experiment, PF elicited by *wh*-questions, which has existential presupposition in Horn's and Zimmerman and Onea's analysis, still deviates from SD.

DeVeaugh-Geiss and colleagues ([6]) attempted to explain the same difference between PF and clefts by resorting to focus projection: since the domain of alternatives of PF (canonical focus for them) is ambiguous while clefts have a clearly designated QUD and a clearly designated domain of alternatives, the former is less optimal for pragmatic enrichment, and thus has weaker exhaustivity. However, in our experiment, all testing sentences were elicited with *wh*-questions, so the domain of alternatives is clearly designated for both PF and SD. If the conversational implicature analysis was to be maintained, an account should be given for this discrepancy.

5 Experiment 3

If the exhaustivity of SD is not conversationally implied, then is it encoded in the presupposition? Boell and Deveaugh-Geiss ([1]) employed a modified "cover-box" design, in which speakers were asked to decide whether knowing that Tom put on a pullover is sufficient to judge the truth/falsity of "it is Tom who put on a pullover." They found that clefts patterned with definite descriptions ("He who put on a pullover is Tom") instead of *only* and plain focus sentences, which supports the presupposition analysis of clefts' exhaustivity.

Specificational Sentence (SS) like (5) takes a similar form as the definite descriptions in Boell and Deveaugh-Geiss ([1])'s study, and was argued to presuppose both existential meaning and exhaustivity. If SD's exhaustivity resembles SS's, then it should be analysed as presupposition as well. As discussed above, presuppositions should project over negation. Therefore, in this experiment, we put SD and SS in projective contexts, and see if their exhaustivity and existential presupposition survive in such contexts.

(6) Women ban li, chidao de (ren) bu *shi* [Zhang Ming]$_F$.
Our class LOC, late DE (person) not SHI Zhang Ming

"In our class, it is not Zhang Ming who was late."
Presupposition: There is someone who was late.
Assertion: Zhang Ming was not late.

Method This experiment is also a felicity judgment task presented as a web-based questionnaire. Forty-seven Mandarin speakers (age: 21-45, mean:27.1) were asked to judge on a scale from 1 to 5 (similar to Experiment 1) the acceptability of David's utterance in a scenario. Each scenario consisted of a short background in written form and a pre-recorded David's utterance as the target sentence, whose written form appeared on the screen after the audio ended.

Each target sentence consisted of two conjuncts: the first was a negated SD or SS sentence which was judged as grammatical by two native Mandarin speakers, and the second part contradicted either the existential meaning (∃Pres) or exhaustivity (EI) of the first conjunct (Tab. 3). The four conditions: sentence type (2 levels: SD/SS) × contradiction type (2 levels: ∃Pres/EI), were tested across 12 sets of sentences, resulting in 48 items. All items were assigned to 4 lists, each of which contained 12 testing and 12 filler items, which was presented in a pseudo-randomised manner.

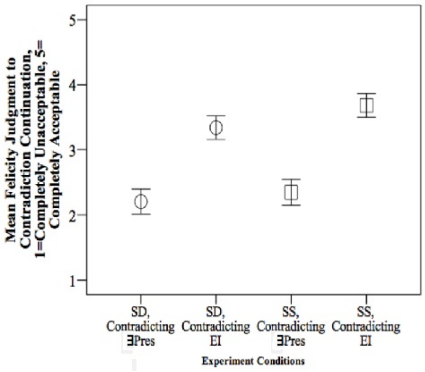
Conjunct 1	Conjunct 2
Women ban li, bu shi [Zhang Ming]$_F$ chidao le. "In our class, it is not Zhang Ming who was late." **(SD)** **∃Pres**: $\exists x[LATE(x)]$; **EI**: $\neg\exists y[LATE(y) \wedge y \neq Zhangming]$	Meiren chidao le. "Nobody was late." **(contradict ∃Pres)**
	Li Jun chidao le. "Li Jun was late." **(contradict EI)**
Women ban li, chidao de (ren) bu shi [Zhang Ming]$_F$. "In our class, the one who was late is not Zhang Ming." **(SS)** **∃Pres**: $\exists x[LATE(x)]$; **EI**: $\neg\exists y[LATE(y) \wedge y \neq Zhangming]$	Meiren chidao le. "Nobody was late." **(contradict ∃Pres)**
	Li Jun chidao le. "Li Jun was late. " **(contradict EI)**

Table 3. Example of four permutations of the target sentence in Experiment 3

Predictions (i) Following the presupposition analysis, the acceptability to the utterances contradicting exhaustivity of SD should pattern with those contradicting the existential meaning; (ii) If the exhaustivity of SD is not presupposed, the acceptability to utterances contradicting exhaustivity of SD should differ from those contradicting the existential meaning.

Results Forty-five completed questionnaires were included in the analysis. The mean acceptability ratings to target sentences in the four conditions are presented in Fig. 3. We can see that for SD, contradicting existential meaning (mean = 2.21) is less acceptable than contradicting exhaustivity (mean = 3.341). This difference is statistically significant tested by one-way ANOVA ($p < 0.0001$), suggesting that the existential meaning of SD can project over negation while exhaustivity cannot. The same pattern was found with SS as well (contradicting existential presupposition mean = 2.35, contradicting exhaustivity mean = 3.68, $p = 0.000$). We can, therefore, conclude that existential meaning is presupposed for SD and SS while exhaustivity is not.

Fig. 3. Contradicting exhaustivity and existential presupposition in SD and SS (means with confidence intervals 95%)



Discussion This experiment shows that contradicting existential meaning renders both negated SD and negated SS unacceptable, indicating that this meaning can project over negation and is presupposed in both constructions. Contradicting exhaustivity, on the other hand, is much more acceptable, suggesting that exhaustivity as $[\neg\exists y[P(y) \wedge y \neq \alpha]$ cannot project over negation. These results fail to support the hypothesis that existential presupposition and exhaustivity are both presuppositions of SD and SS.

6 General Discussion

Our experiments presented empirical evidence challenging the assertion, presupposition and conversational implicature hypotheses of the exhaustivity of SD clefts. We would then attempt to propose an analysis for SD that could account for the above results.

A pattern specific to Mandarin is that SD reflects certain definite effect. For SD, the cleft focus selects a wide range of elements, such as quantifiers and predicates (c.f. the quantifier and predicate constraints of English *it*-clefts, see [10]), and yet exhaustivity arises only when the cleft focus is a referring expression:

(7) Shi [yixie ren]$_F$ fan le cuowu. ***(Non-exhaustive)***
SHI some people make ASP a mistake

Intended: '[Some people]$_F$ made a mistake ($^{??}$not other people).'

(8) Zhangsan shi [qiaoqiaode]$_F$ liuzou de. ***(Non-exhaustive)***
Zhangsan SHI quietly slipped away DE

Intended: 'Zhangsan slipped away [quietly]$_F$ ($^{??}$not in another manner).'

These observations direct us back to an epiphenomenal proposal in line with Percus ([19]) and Hedberg ([11]), in which exhaustivity of clefts is closely related to the maximality of a covert iota-operator introduced by the referring expression as cleft focus. In a SD cleft where the cleft focus is an e type referring expression, the copula *shi* denotes an equative relation $\lambda x \lambda y[y = x]$. Since the cleft focus on one side of the equation is of type e while the cleft clause on the other side is of type $< e, t >$, the latter must be shifted to type e to avoid type mismatch, which is realized by the iota-operator. After type-shifting, the cleft clause is related to a definite expression, carrying a maximality presupposition. The semantics of (1) could now be re-written as (9):

(9) Shi [Xiaogao he Xiaopang]$_F$ chidao le.
SHI Xiaogao and Xiaopang late ASP

'It is Xiaogao and Xiaopang who were late.'
Presupposition: $\exists x.x \in {}^*[\![\text{LATE}]\!] \wedge \forall y.y \in {}^*[\![\text{LATE}]\!] \rightarrow y \leq x$
(x is the maximal entity that were late)
Assertion: $x = \mathbf{MAX}({}^*[\![\text{Xiaogao and Xiaopang}]\!])$
(The maximal entity in "Xiaogao and Xiaopang" equals x)

In presupposition, x is the maximal member in the denotation of the cleft clause under closure; in assertion, this x equals to the maximal member in the denotation of the cleft focus under closure. Combining the presupposition with assertion, exhaustivity is derived: in (9), the maximal member in the set of all individuals being late is Xiaogao and Xiaopang, inferring that no one else was late. [4]

[4] Previously, scholars ([21], [3]) noticed the following pattern:

(1) $^{??}$ It wasn't Alice who laughed, it was Alice and Bob. ([21, p.15])

Note that Hedberg ([11]) assumes that the two pivotal components of all clefts, cleft pronoun *it* and the cleft clause, together form a definite expression. In this way, her proposal predicts that all cleft sentences are exhaustive. We, on the other hand, do not make such an assumption. By attributing exhaustivity to the combination of referring expression, equative *shi* and iota-operator, we have detached exhaustivity from cleft structure. In this way, SD sentences without iota-operator or canonical sentences without equative *shi* are not exhaustive; in adverbial, quantification or indefinite focused SD sentences, the denotation of *shi* is vacuous instead of an equative relation, these SD sentences are exhaustive either.

For negated SD clefts in Experiment 3, though presupposition projects over negation, the asserted content is negated, so that x no longer equals to the maximal member of cleft focus. As a result, the maximal member of cleft focus does not equal to the maximal member of cleft clause; exhaustivity is thus effaced. Moreover, being derived from a presupposed iota-operator and an asserted equative relation, exhaustivity in SD cleft is semantic in nature but not a prototypical assertion like the exhaustivity of ZY, which explains the different levels of acceptance to exhaustivity among ZY, SD and PF in Experiment 1 and the low level of cancelability of SD's exhaustivity in Experiment 2.

References

1. Boell, A.C., Deveaugh-Geiss., J.P.: Empirical insights on the exhaustivity inference in *it*-clefts (2015), presentation at Experimental Approaches to Semantics Workshop.
2. Büring, D.: Conditional exhaustivity presuppositions in clefts (and defi- nites) (2011), ms. ZAS/University of Vienna

(2) $^{??}$ It wasn't Fred she invited. She invited Fred and Gordon. ([3, p.2])

(3) $^{??}$ It wasn't Fred and George she invited. She invited Fred.

However, the pattern in Mandarin is not so clear; most of the native speakers we have consulted found (4) acceptable. If Mandarin resembles English in this aspect, one possible way to fix this problem may be to add $[\exists z.z \in {}^*[\![\text{Xiaogao and Xiaopang}]\!] \wedge z \leq x]$ to assertion as in (4):

(4) Bu shi Xiaogao he Xiaopang chidao le; shi Xiaogao chidao le.
Not SHI Xiaogao and Xiaopang late ASP; SHI Xiaogao late ASP
'It is not Xiaogao and Xiaopang who were late; Xiaogao were late'
Assertion: $\neg[\exists z.z \in {}^*[\![\text{Xiaogao and Xiaopang}]\!] \wedge z \leq x]$;
$x \neq \mathbf{MAX}({}^*[\![\text{Xiaogao and Xiaopang}]\!])$

By adding this component, all entities in the domain of individuals that consist Xiaogao or Xiaopang as a proper part (e.g. [Xiaogao and Xiaoming], [Xiaogao, Xiaopang and Xiaoming]) are not entities that were late.

It is also likely that the varied judgments on such sentences are due to pragmatic reasons; we will investigate this problem in future research.

3. Büring, D., Križ, M.: Exhaustivity and homogeneity presupposition in clefts (and definites). Semantics and Pragmatics 6(6), 1–29 (2013)
4. Byram-Washburn, M., Kaiser, E., Zubizarreta, M.L.: The English *It*-Cleft: No Need to Get Exhausted. Ph.D. thesis, University of Southern California (2013)
5. Destruel, E., Velleman, D., Onea, E., Bumford, D., Xue, J., Beaver, D.: A cross-linguistic study of the non-at-issueness of exhaustive inferences. In: Schwarz, F. (ed.) Experimental Perspectives on Presuppositions Experimental Perspectives on Presuppositions, pp. 135–156. Springer (2015)
6. DeVeaugh-Geiss, J.P., Zimmermann, M., Onea, E., Boell, A.: Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. In: D'Antonio, S., Moroney, M., Little, C.R. (eds.) The 25th Semantics and Linguistic Theory Conference (2015)
7. É Kiss, K.: Identificational focus vs. information focus. Language 74(2), 245–273 (1998)
8. Grice, P.: Studies in the Way of Words. Harvard University Press, Boston (1989)
9. Halvorsen, P.K.: The Syntax and Semantics of Cleft Constructions. Ph.D. thesis, University of Texas, Austin (1978)
10. Hedberg, N.: Discourse Pragmatics and Cleft Sentences in English. Ph.D. thesis, University of Minnesota (1990)
11. Hedberg, N.: The referential status of clefts. Language 76, 891–920 (2000)
12. Horn, L.: Information structure and the landscape of (non-)atissue meaning. In: Féry, C., Ishihara, S. (eds.) The Oxford Handbook of Information Structure,. Oxford University Press. (to appear)
13. Horn, L.: Exhaustiveness and the semantics of clefts. In: Burke, V., Pustejovsky, J. (eds.) Proceedings of the Eleventh Annual Meeting of the North Eastern Linguistics Society (NELS) (1981)
14. Karttunen, L.: Presuppositions and compound sentences. Linguistic Inquiry 4(2), 169–193 (1973)
15. Karttunen, L., Peters, S.: Conventional implicature. In: Oh, C.K., Dinneen, D. (eds.) Syntax and Semantics, vol. 11: Presupposition. Academic Press (1979)
16. Lee, H.c.: On Chinese Focus and Cleft Constructions. Ph.D. thesis, National Tsing Hua University, Hsinchu (2005)
17. Mayol, L., Castroviejo, E.: How to cancel an implicature. Journal of Pragmatics 50, 84–104 (2013)
18. Onea, E., Beaver, D.: Hungarian focus is not exhausted. In: Cormany, E., Ito, S., Lutz, D. (eds.) Proceedings of the 19th Semantics and Linguistic Theory (SALT) (2009)
19. Percus, O.: Prying open the cleft. In: Kusumoto, K. (ed.) the 27th Annual Meeting of the North-East Linguistics Society (NELS). p. 337ñ351 (1997)
20. Teng, S.h.: Remarks on cleft sentences in Chinese. Journal of Chinese Linguistics 7, 101–113 (1979)
21. Velleman, D.B., Beaver, D., Destruel, E., Bumford, D., Onea, E., Coppock, L.: *It*-clefts are it (inquiry terminating) constructions. In: Chereches, A. (ed.) Proceedings of the 22th Semantics and Linguistic Theory (SALT). pp. 441–420 (2012)
22. Zimmermann, M., Onea, E.: Focus marking and focus interpretation. Lingua 121(11), 1651–1670 (2011)

Hilbert-style Lambek calculus with two divisions

Valentina Lugovaya and Anastasiia Ryzhova

Lomonosov Moscow State University

v.lugovaya8@gmail.com, nas-ryzhova@yandex.ru

Abstract. This paper is concerned with the Lambek calculus of syntactic types (introduced in 1958 by J. Lambek). There exist two variants of the Lambek calculus: sequential L, and non-sequential (Hilbert-style) $\mathrm{L_H}$. Their product-free fragments with two divisions are considered here. $\mathrm{L}(/,\backslash)$ results naturally due to the subformula property. We construct $\mathrm{L_H}(/,\backslash)$ and prove its equivalence to the $\mathrm{L}(/,\backslash)$ with the help of semantic methods going back to the works of W. Buszkowski.

In 1958 J. Lambek introduced a calculus for deriving reduction laws of syntactic types (see [1]). The Lambek calculus exists in two variants: sequential and non-sequential (Hilbert-style).

First we consider L — the sequential variant of the Lambek calculus. Its types are built from the primitive types $(p_1, p_2, ...)$ using three binary connectives: $\cdot$ (multiplication), / (right division) and \ (left division). Types are denoted by capital Latin letters. Capital Greek letters denote finite (possibly empty) linearly ordered sequences of types. Sequents of L are expressions of the form $\Pi \to C$, where Π is non-empty. L is specified by axioms of the form $p_i \to p_i$ and the following derivation rules:

$$\frac{A\,\Pi \to B}{\Pi \to A \backslash B}\ (\to \backslash),\ \text{where } \Pi \text{ is non-empty} \qquad \frac{\Pi \to A \quad \Gamma\, B\, \Delta \to C}{\Gamma\, \Pi\, (A \backslash B)\, \Delta \to C}\ (\backslash \to)$$

$$\frac{\Pi\, A \to B}{\Pi \to B / A}\ (\to /),\ \text{where } \Pi \text{ is non-empty} \qquad \frac{\Pi \to A \quad \Gamma\, B\, \Delta \to C}{\Gamma\, (B / A)\, \Pi\, \Delta \to C}\ (/ \to)$$

$$\frac{\Gamma \to A \quad \Delta \to B}{\Gamma\, \Delta \to A \cdot B}\ (\to \cdot) \qquad \frac{\Gamma\, A\, B\, \Delta \to C}{\Gamma\, (A \cdot B)\, \Delta \to C}\ (\cdot \to)$$

$$\frac{\Pi \to A \quad \Gamma\, A\, \Delta \to C}{\Gamma\, \Pi\, \Delta \to C}\ (\text{cut})$$

As it has been proved in [1], the last rule (cut) is eliminable.

As for the Hilbert-style Lambek calculus $\mathrm{L_H}$, its formulas are expressions of the form $A \to B$, where A and B are types. $\mathrm{L_H}$ is specified by axioms of the form $A \to A$, $(A \cdot B) \cdot C \to A \cdot (B \cdot C)$, $A \cdot (B \cdot C) \to (A \cdot B) \cdot C$, and the following derivation rules:

$$\frac{A \cdot C \to B}{C \to A \backslash B} \qquad \frac{C \cdot A \to B}{C \to B / A} \qquad \frac{C \to A \backslash B}{A \cdot C \to B} \qquad \frac{C \to B / A}{C \cdot A \to B} \qquad \frac{A \to B \quad B \to C}{A \to C}$$

The equivalence of the two variants of the Lambek calculus has been proved by J. Lambek in [1]:

Theorem 1 (Lambek '58). $\mathrm{L_H} \vdash A \to B \Longleftrightarrow \mathrm{L} \vdash A \to B.$

If we restrict the set of connectives in L, the calculus results naturally due to the subformula property (it is just necessary to exclude the rules containing the removed connectives). But it is not the case for the non-sequential variant $\mathrm{L_H}$. Thus, the construction of the fragments of $\mathrm{L_H}$ with restricted sets of connectives is nontrivial. For the one-division fragment $\mathrm{L}(\backslash)$ this was done in [5] (unfortunately, this paper has never been published). In our paper we construct $\mathrm{L_H}(/, \backslash)$ and prove its equivalence to $\mathrm{L}(/, \backslash)$. The choice of this very fragment is quite natural since "real" grammars built for linguistic applications do not usually involve multiplication. Thus, the study of this fragment appears to be interesting especially since the Hilbert variant for it has not been hitherto known.

We define the calculus $\mathrm{L_H}(/, \backslash)$ as follows:

Definition 1. $\mathrm{L_H}(/, \backslash)$ *is specified by the following axioms:*
(1) $A \to A$ (2) $B \,\backslash\, C \to (A \,\backslash\, B) \,\backslash (A \,\backslash\, C)$ (3) $A \,\backslash (B \,/\, C) \leftrightarrow (A \,\backslash\, B) \,/\, C$ *and the following derivation rules:*

$$\frac{A \to B \qquad B \to C}{A \to C}\ (1) \qquad\qquad \frac{A \to B \qquad C \to D}{B \,\backslash\, C \to A \,\backslash\, D}\ (2)$$

$$\frac{A \to B \,/\, C}{C \to A \,\backslash\, B}\ (3) \qquad\qquad \frac{A \to B \,\backslash\, C}{B \to C \,/\, A}\ (4)$$

Axiom (2) is actually a well-known principle called *Geach rule.* It was introduced by Geach in [4] (but in a slightly different form) and many times used for alternative axiomatizations of the Lambek calculus in papers on categorial grammar, e.g., [5], [6], [7], [8], [9] etc.

It is significant that all the equivalence proofs in [4], [5], [6] and [7] were purely syntactic, whereas our proof is based on semantic methods.

And now we prove the equivalence of $\mathrm{L_H}(/, \backslash)$, constructed above, and $\mathrm{L}(/, \backslash)$.

Theorem 2. $\mathrm{L_H}(/, \backslash) \vdash A \to B \Longleftrightarrow \mathrm{L}(/, \backslash) \vdash A \to B.$

Proof. 1) $\boxed{\Rightarrow}$ This is an easy, syntactic part of the proof.

It is necessary to show that axioms (2) and (3) and all the rules of the calculus $\mathrm{L_H}(/, \backslash)$ are derivable in the sequential Lambek calculus $\mathrm{L}(/, \backslash)$.

The Geach rule and axiom (3) are known to be derivable in L, and therefore in $\mathrm{L}(/, \backslash)$. Rule (1) is just a particular case of (cut). Rule (2) is the monotonicity rule, admissible in $\mathrm{L}(/, \backslash)$ due to [1].

Adduce the derivations of rules (3) and (4) in $\mathrm{L}(/, \backslash)$:

$$\dfrac{\dfrac{A \to B \,/\, C \qquad \dfrac{C \to C \qquad B \to B}{B \,/\, C \quad C \to B}\ (/ \to)}{A \quad C \to B}\ (\text{cut})}{B \to C \,/\, A}\ (\to \backslash) \qquad\qquad \dfrac{\dfrac{A \to B \,\backslash\, C \qquad \dfrac{B \to B \qquad C \to C}{B \quad B \,\backslash\, C \to C}\ (\backslash \to)}{B \quad A \to C}\ (\text{cut})}{C \to A \,\backslash\, B}\ (\to /)$$

Therefore $\mathrm{L_H}(/, \backslash)$ is correct with respect to $\mathrm{L}(/, \backslash)$.

2) $\boxed{\Leftarrow}$ Since the sequential calculus $\mathrm{L}(/, \backslash)$ is more convenient for derivations, the first part of the proof involved syntactic methods. But it is not the case for the Hilbert-style variant $\mathrm{L_H}(/, \backslash)$, that is the reason why in this part it is appropriate to use semantic proof methods. Note that syntactic method still could be applied to the proof of this direction (in the spirit of [5]), but the proof will be not that elegant.

Let us define the natural interpretation of the Lambek calculus.

Definition 2. *An L-model (a language model, a model on the subsets of a free semigroup) is a structure $\mathcal{M} = \langle \Sigma, w \rangle$, where Σ is a finite or countable alphabet and $w\colon \mathrm{Tp} \to \mathcal{P}(\Sigma^+)$ is a function from the set of the types of the Lambek calculus into the set of the languages on the alphabet Σ, satisfying the following conditions: $w(A{\cdot}B) = w(A){\cdot}w(B), w(A \setminus B) = w(A) \setminus w(B), w(A \,/\, B) = w(A) \,/\, w(B)$ for any types A and B. A formula $A \to B$ is true in $\mathcal{M}$ iff $w(A) \subseteq w(B)$.*

The operations on languages are defined as follows:

Definition 3. *Let A and B be languages on the alphabet Σ. Then*
$A \,/\, B \rightleftharpoons \{u \in \Sigma^+ \mid \forall v \in B \;\; uv \in A\}$
$B \setminus A \rightleftharpoons \{u \in \Sigma^+ \mid \forall v \in B \;\; vu \in A\}$
$A \cdot B \rightleftharpoons \{uv \in \Sigma^+ \mid u \in A, \; v \in B\}$

According to [3], L is sound and complete with respect to this standard interpretation of the Lambek calculus connectives as operations on formal languages. We will use an analogous theorem for $\mathrm{L}(/, \backslash)$ proved by W. Buszkowski in [2].

Theorem 3 (Buszkowski '82). $\mathrm{L}(/, \backslash) \vdash A \to B \iff \forall \Sigma \, \forall w \;\; w(A) \subseteq w(B)$.

In order to prove this theorem W. Buszkowski has constructed the canonical model $\langle \Sigma_0, w_0 \rangle$, where $\Sigma_0 = \mathrm{Tp}(/, \backslash)$, $w_0(A) = \{\Gamma \in \Sigma_0^+ \mid \mathrm{L}(/, \backslash) \vdash \Gamma \to A\}$. This model appears to be universal, that is, it makes true exactly those formulas that are derivable in $\mathrm{L}(/, \backslash)$: $\mathrm{L}(/, \backslash) \vdash A \to B \iff w_0(A) \subseteq w_0(B)$.

We will prove L-completeness of $\mathrm{L_H}(/, \backslash)$ using the same method.

Let $\mathrm{L}(/, \backslash) \vdash A \to B$, then for all $\mathcal{M} = \langle \Sigma, w \rangle \;\; w(A) \subseteq w(B)$, according to W. Buszkowski. Following him, we will define a slightly different model for the Hilbert-style calculus: $\Sigma_0 \rightleftharpoons \mathrm{Tp}(/, \backslash)$ and $w_0(A) \rightleftharpoons \{\, B_1 B_2 \ldots B_n \mid \mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \setminus \ldots \setminus (B_2 \setminus (B_1 \setminus A)) \,\}$ for all $A \in \mathrm{Tp}(/, \backslash)$.

First, let us check that $\langle \Sigma_0, w_0 \rangle$ is really an L-model, that is:

1. $w_0(C \,/\, D) = w_0(C) \,/\, w_0(D)$
2. $w_0(C \setminus D) = w_0(C) \setminus w_0(D)$

1. $\boxed{\subseteq}$ Let $B_1 \ldots B_n \in w_0(C \,/\, D)$, $E_1 \ldots E_k \in w_0(D)$. Let us prove that $B_1 \ldots B_n E_1 \ldots E_k \in w_0(C)$.

$B_1 \ldots B_n \in w_0(C \,/\, D) \Rightarrow\ \mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash (B_1 \backslash (C \,/\, D))$.

$E_1 \ldots E_k \in w_0(D) \Rightarrow\ \mathrm{L_H}(/, \backslash) \vdash E_k \to E_{k-1} \backslash \ldots \backslash E_1 \backslash D$.

Our goal is to show that $\mathrm{L_H}(/, \backslash) \vdash E_k \to E_{k-1} \backslash \ldots \backslash E_1 \backslash B_n \backslash \ldots \backslash B_1 \backslash C$. Here is the derivation of this sequence. The sequence $B_{n-1} \backslash \ldots \backslash (B_1 \backslash (C \,/\, D)) \to (B_{n-1} \backslash \ldots \backslash (B_1 \backslash C)) \,/\, D$ is derivable in $\mathrm{L_H}(/, \backslash)$ by induction on n applying axiom (3) and rule (2). Next,

$$\dfrac{\dfrac{B_n \to B_{n-1} \backslash \ldots \backslash (B_1 \backslash (C \,/\, D)) \qquad B_{n-1} \backslash \ldots \backslash (B_1 \backslash (C \,/\, D)) \to (B_{n-1} \backslash \ldots \backslash (B_1 \backslash C)) \,/\, D}{B_n \to (B_{n-1} \backslash \ldots \backslash (B_1 \backslash C)) \,/\, D}\,(1)}{D \to B_n \backslash \ldots \backslash B_1 \backslash C}\,(3)$$

$$\dfrac{E_k \to E_{k-1} \backslash \ldots \backslash E_1 \backslash D \qquad \dfrac{E_{k-1} \backslash \ldots \backslash E_1 \to E_{k-1} \backslash \ldots \backslash E_1 \qquad D \to B_n \backslash \ldots \backslash B_1 \backslash C}{E_{k-1} \backslash \ldots \backslash E_1 \backslash D \to E_{k-1} \backslash \ldots \backslash E_1 \backslash B_n \backslash \ldots \backslash B_1 \backslash C}\,(2)}{E_k \to E_{k-1} \backslash \ldots \backslash E_1 \backslash B_n \backslash \ldots \backslash B_1 \backslash C}\,(1)$$

$\boxed{\supseteq}$ Let $B_1 \ldots B_n \in w_0(C) \,/\, w_0(D)$. Let us prove that $B_1 \ldots B_n \in w_0(C \,/\, D)$, i.e. $\mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash B_1 \backslash (C \,/\, D)$.

For all $\varGamma \in w_0(D)$ we have $B_1 \ldots B_n \varGamma \in w_0(C)$. Let $\varGamma = D$, then $B_1 \ldots B_n D \in w_0(C)$, i.e. $\mathrm{L_H}(/, \backslash) \vdash D \to B_n \backslash \ldots \backslash B_1 \backslash C$. Let us derive the required formula:

$$\dfrac{\dfrac{D \to B_n \backslash \ldots \backslash B_1 \backslash C}{B_n \to (B_{n-1} \backslash \ldots \backslash (B_1 \backslash C)) \,/\, D}\,(4) \qquad (B_{n-1} \backslash \ldots \backslash (B_1 \backslash C)) \,/\, D \to B_{n-1} \backslash \ldots \backslash B_1 \backslash (C \,/\, D)}{B_n \to B_{n-1} \backslash \ldots \backslash B_1 \backslash (C \,/\, D)}\,(1)$$

2. $\boxed{\subseteq}$ Let $B_1 \ldots B_n \in w_0(C \backslash D)$, $E_1 \ldots E_k \in w_0(C)$. Let us prove that $E_1 \ldots E_k B_1 \ldots B_n \in w_0(D)$.

$B_1 \ldots B_n \in w_0(C \backslash D) \Rightarrow\ \mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash (B_1 \backslash (C \backslash D))$.

$E_1 \ldots E_k \in w_0(C) \Rightarrow\ \mathrm{L_H}(/, \backslash) \vdash E_k \to E_{k-1} \backslash \ldots \backslash E_1 \backslash C$.

We need to obtain that $\mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash B_1 \backslash E_k \backslash \ldots \backslash E_1 \backslash D$. Here is the derivation of this sequent:

$$\dfrac{E_k \to E_{k-1} \backslash \ldots \backslash E_1 \backslash C \qquad E_{k-1} \backslash \ldots \backslash E_1 \backslash D \to E_{k-1} \backslash \ldots \backslash E_1 \backslash D}{(E_{k-1} \backslash \ldots \backslash E_1 \backslash C) \backslash (E_{k-1} \backslash \ldots \backslash E_1 \backslash D) \to E_k \backslash (E_{k-1} \backslash \ldots \backslash E_1 \backslash D)}\,(2)$$

Using axiom (2): $C \backslash D \to (E_{k-1} \backslash \ldots \backslash E_1 \backslash C) \backslash (E_{k-1} \backslash \ldots \backslash E_1 \backslash D)$ and rule (1) we obtain: $C \backslash D \to E_k \backslash \ldots \backslash E_1 \backslash D$.

Further,

$$\dfrac{B_n \to B_{n-1} \backslash \ldots \backslash B_1 \backslash (C \backslash D) \qquad \dfrac{B_{n-1} \backslash \ldots \backslash B_1 \to B_{n-1} \backslash \ldots \backslash B_1 \qquad C \backslash D \to E_k \backslash \ldots \backslash E_1 \backslash D}{B_{n-1} \backslash \ldots \backslash B_1 \backslash (C \backslash D) \to B_{n-1} \backslash \ldots \backslash B_1 \backslash E_k \backslash \ldots \backslash E_1 \backslash D}\,(2)}{B_n \to B_{n-1} \backslash \ldots \backslash B_1 \backslash E_k \backslash \ldots \backslash E_1 \backslash D}\,(1)$$

$\boxed{\supseteq}$ Let $B_1 \ldots B_n \in w_0(C) \backslash w_0(D)$. Let us prove that $B_1 \ldots B_n \in w_0(C \backslash D)$, i.e. $\mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash B_1 \backslash (C \backslash D)$.

For all $\varGamma \in w_0(C)$ we have $\varGamma B_1 \ldots B_n \in w_0(D)$. Let $\varGamma = C$, then $C B_1 \ldots B_n \in w_0(D)$, i.e. $\mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash B_1 \backslash C \backslash D$.

Finally, since $\mathrm{L}(/, \backslash) \vdash A \to B$ and $\langle \Sigma_0, w_0 \rangle$ is an L-model, $w_0(A) \subseteq w_0(B)$, and it remains to show that if $w_0(A) \subseteq w_0(B)$, then $\mathrm{L_H}(/, \backslash) \vdash A \to B$. Since $A \in w_0(A) \subseteq w_0(B)$, hence $\mathrm{L_H}(/, \backslash) \vdash A \to B$ (by the construction of w_0). Q.E.D.

□

Now that we have proved the equivalence of $\mathrm{L}(/, \backslash)$ and $\mathrm{L_H}(/, \backslash)$, we see that actually Buszkowski's model, where $w(A) = \{\, \Gamma \in \mathrm{Tp}(/, \backslash)^+ \mid \mathrm{L}(/, \backslash) \vdash \Gamma \to A \,\}$, and our model, where $w(A) = \{\, B_1 B_2 \ldots B_n \mid \mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash (B_2 \backslash (B_1 \backslash A)) \,\}$, are equivalent, since
$\mathrm{L_H}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash (B_2 \backslash (B_1 \backslash A)) \iff$
$\iff \mathrm{L}(/, \backslash) \vdash B_n \to B_{n-1} \backslash \ldots \backslash (B_2 \backslash (B_1 \backslash A)) \iff$
$\iff \mathrm{L}(/, \backslash) \vdash B_1 \ldots B_n \to A$.

Our argument also works for $\mathrm{L}(\backslash)$ and $\mathrm{L_H}(\backslash)$, thus giving a simpler proof of Savateev's result [5]. The advantage of our construction is that we can describe Savateev's calculus $\mathrm{L_H}(\backslash)$ just as a fragment of our calculus ($\mathrm{L_H}(\backslash)$ is obtained from $\mathrm{L_H}(/, \backslash)$ presented above by removing axiom (3) and rules (3) and (4)).

Acknowledgements

We are very grateful to Assistant Professor Stepan Kuznetsov for guiding us into the subject and his constant attention to our work.

References

1. Lambek, J.: The mathematics of sentence structure. American Mathematical Monthly 65, 3, 154–170 (1958)
2. Buszkowski, W.: Compatibility of a categorial grammar with an associated category system. Zeitschrift für mathematische Logik und Grundlagen der Mathematik 28, 229–238 (1982)
3. Pentus, M.: Models for the Lambek calculus. Annals of Pure and Applied Logic 75, 1–2, 179–213 (1995)
4. Geach, P. T.: A program for syntax. In D. Davidson and G. Harman, Semantics of Natural Language, Reidel, Dordrecht 483–497 (1972)
5. Savateev, Y.: Variants of the Lambek calculus with one division. Unpublished manuscript (in Russian). Moscow State University (2004)
6. Cohen, J. M.: The equivalence of two concepts of categorial grammar. Information and Control 10, 475–484 (1967)
7. Zielonka, W.: Axiomatisability of Ajdukiewicz-Lambek Calculus by Means of Cancellation Schemes. Zeitschrift für mathematische Logik und Grundlagen der Mathematik 27, 215–224 (1981)
8. Steedman, M.: Categorial Grammar. Technical Reports (CIS). Paper 466 (1992)
9. Humberstone, L.: Geach's Categorial Grammar. Linguistics and Philosophy 28, 3, 281–317 (2005)

Knowledge reports without truth

Deniz Özyıldız

University of Massachusetts, Amherst

1 Introduction

I present novel[1] data from Turkish attitude reports introduced by factive predicates where the potential falsity of the proposition that expresses the attitude content does not trigger infelicity: Non-factive reports introduced by factive predicates. Emphasis is placed on "non-factive knowledge."

In (1), *bil-*, an attitude predicate corresponding to the English *know* is used, yet the sentence is felicitous in a context that makes the proposition expressed by the embedded clause false, that is, where Bernie did not win the election. In fact, sentences like (1) often give rise to the inference that the attitude holder's belief is mistaken.

(1) **Context**: Trump won the election, but…
Tunç [Bernie kazan-dı diye] biliyor.
Tunç Bernie won-PST.3S *diye* knows
Tunç *thinks* (lit. #*knows*) that Bernie won.

In (2), the attitude predicate is used with a nominalized embedded clause. The sentence is not felicitous when the proposition expressed by the embedded clause is false.

(2) **Context**: Trump won the election, but…
# Tunç [Bernie-nin kazan-dığ-ı-nı] biliyor.
Tunç Bernie-GEN win-NMZ-3S-ACC knows
# Tunç knows that Bernie won.

This paper explores this factivity alternation, lays out the analytical challenge that it raises, and proposes to give sentence (1) roughly the semantics in (3):

(3) Tunç knows something, which gives rise to the belief that Bernie won.
$\not\leadsto$ Bernie won.[2]

[1] Translations of examples by Şener, [3], suggest that at least some Turkish linguists are aware of the facts: In (i), *bil-* is translated by *think* instead of *know*.

(i) Pelin [sen Timbuktu-ya git-ti-n diye] bil-iyor-muş.
Pelin you Timbuktu-DAT go-PST-2S *diye* know-PRES.EVID
Pelin *thinks* (lit. *knows*) that you went to Timbuktu. Adapted from [3], ex. (4)

Native speakers use this construction productively.

[2] Squiggly arrows introduce presupposed content.

The knowledge component is directly contributed by the matrix predicate. The function of *diye* is to introduce a secondary belief component as well as its propositional content. The matrix attitude predicate's internal argument position is saturated by a pronoun, valued by the assignment function. The pronoun picks out a relevant set of facts that motivates the belief introduced by *diye*. (Perhaps the attitude holder has watched a false news report.) For concreteness, I assume that there is a causal link between what is known and what is believed, which might be contributed by the way *diye* composes with the matrix attitude predicate.

Given that the proposition introduced by *diye*, in (1), is a *belief* proposition, it is not presupposed and the sentence is not factive. Sentence (2), on the other hand, has the nominalized clause directly saturating the matrix predicate's internal argument [4]:

(4) Tunç knows (the fact) that Bernie won. ⇝ Bernie won.

The embedded proposition is a *knowledge* proposition, and the sentence is factive.

Before moving on to the next section, I would like to make a few comments on the factivity alternation illustrated by the contrast between (1) and (2). There is a general consensus in the literature that the presupposition associated with a certain class of factives including *know*, is "weak," in the sense that it is easily suspended. The original observations seem to be from Karttunnen [9] and the Kiparskys [12].

(5) a. I don't know that this isn't our car. [12]
⇝̸This isn't our car.
b. Did you discover that you had not told the truth? [9]
⇝̸You had not told the truth.

The factivity alternation discussed here seems to be distinct from the effects in (5). In particular, non-factive uses of factive predicates in English seem to cooccur with some sentential operator like negation or question (other examples in the literature include conditionals, modals and focus [21]), and depend on contextual factors [20] and indexicality. Such factors are not required to generate the Turkish contrast under scrutiny, which mainly appears to be conditioned by the syntax of the embedded clause.

Three hypotheses can be formulated on the basis of this factivity alternation. First, the *homophony hypothesis* states that Turkish has two homophonous *bil-* predicates, one lexically factive, and the other not. This duplication is required for every predicate that participates in the alternation. (I do not dwell much upon this hypothesis.)

The two other options are based on deriving the alternation from a unified semantics for attitude predicates. The *external factivity hypothesis* states that *bil-* and other 'factives' have a single, non-factive lexical entry. The account based on it generates factivity compositionally. This line of thinking is instantiated in [2, 16, 18], who argue that attitude predicates do not directly impose conditions (like a truth presupposition) on propositions. Finally, the *lexical factivity hypothesis* states that *bil-* and other factives have a single, factive lexical entry. The account based on it derives non-factive reports by introducing a device for suspending the factive presupposition. These two accounts are difficult to tease apart. Although I am unable to dismiss the external factivity hypothesis decisively, I argue that the lexical factivity hypothesis is better fit to handle the Turkish alternation, as it is able to capture the patterns in the data, without placing a load on the lexicon and on selectional properties.

As a final remark, I take the ‘justified true belief’ definition of knowledge to be an accurate working hypothesis for the meaning of *factive* know. I show that, with *non-factive* know, although the belief proposition need not be true, it nevertheless requires justification. This makes non-factive knowledge reports different, on the one hand, from factive ones (which require truth), and, on the other, from neutral belief reports (which do not require justification). I do not intend to make claims here about the *definition* of knowledge, say, from a philosophical standpoint. But non-factive knowledge reports should prove to be an interesting case study for the epistemologist.

2 The factivity alternation

This section first compares non-factive knowledge reports with factive ones and with plain belief reports. The result is that something, namely justification, is retained of knowledge. The second subsection generalizes this result to other attitude predicates, and argues that *diye*’s function is to introduce a secondary speech/attitude predicate, along with its content, that is scopally independent from the matrix predicate.

2.1 A case study: Non-factive know

In (1), *diye bil-* construction was translated as *thinks*, as opposed to *bil-* with a nominalization, translated as *knows*. Overall an attitude report with factive know is (often) only felicitous if the belief proposition is both justified and true. An attitude report with non-factive know requires the belief to be justified, but not that it be true. A neutral belief report requires neither justification nor truth. This is summarized in (6):

(6)

		requirement		
		belief	*justification*	*truth*
att. pred.	*factive know*	yes	yes	yes
	non-factive know	yes	yes	no
	think	yes	no	no

To motivate this result, I report in table (9), the felicity of the sentences in (7) across four conditions crossing justification of the belief (J, ¬J) and truth of the belief proposition (T, ¬T). Different conditions are obtained by minimally manipulating the context in which the sentences in (7) are judged. These contexts are provided in (8). The judgments reported in (9) are based on my own native intuition. The present set up could serve to run a wider scale controlled felicity judgment experiment, as an anonymous reviewer suggests. I must leave this for further research.

The content of the belief is kept constant: that *Bernie won*. The nature of the attitude varies: A tensed clause clause introduced by *diye* composes with *bil-*, in (7a), to give rise to non-factive know. A nominalization composes with *bil-* in (7b) for factive know, and a nominalization[3] with *düşün-*, 'think,' in (7c) for neutral belief.

(7) a. Tunç [Bernie kazan-dı diye] biliyor.
Tunç Bernie win-PST.3S *diye* knows
Tunç *thinks* (lit. *knows*) that Bernie won.

b. Tunç [Bernie-nin kazan-dığ-ı-nı] biliyor.
Tunç Bernie-GEN win-NMZ-3S-ACC knows
Tunç knows that Bernie won.

c. Tunç [Bernie-nin kazandığ-ı-nı] düşünüyor.
Tunç Bernie-GEN win-NMZ-3S-ACC thinks
Tunç thinks that Bernie won.

Suppose the overall context in (8), and the four conditions in (a-d).

(8) **Overall context for (7)**: Tunç is in solitary confinement when the US presidential election happens. He has no access to the news, and his guards do not communicate with him. He gets out after the elections. Somebody teases him: "So, who won?" Tunç, who is a fervent Bernie supporter, says "Bernie won."

Conditions:

a. Tunç has no information. Trump won. $\neg J, \neg T$

b. Tunç has no information. Bernie won. $\neg J, T$

c. Tunç overheard some talk about Bernie's victory during his confinement. This was a prank! Trump won. $J, \neg T$

d. Tunç overheard some talk about Bernie's victory during his confinement. This was not a prank. Bernie won. J, T

The first pattern to note in table (9) is that justification[4] licenses the use of *bil-*, 'know,' in general. This is seen in the contrast between the first three rows and the last three rows in the table. 'Think' is felicitous in the absence of justification, regardless of whether the belief proposition is true or false. The second observation is that factive 'know' is not licensed by justification alone, but that it requires truth as well. The final observation is that if factive 'know' is licensed, non-factive 'know' and 'think' sound odd, and not maximally collaborative. This is indicated by parentheses around checkmarks (✓) in the lower right quadrant. This effect seems to be pragmatic [19].

[3] I fail to detect any meaning difference between sentences where a tensed clause under 'think,' compared to those with a nominalization. Further research is required here.

[4] For the importance of justification in ensuring the felicity of knowledge ascription, see Kratzer's [14] discussion of a Bertrand Russell example. In these examples, the kind of justification that suffices to license both factive and non-factive know is weak (hearsay). Problematic justification, such as Gettier cases are not discussed [5]. What kind of evidence is 'good enough' to license knowledge ascriptions with *dye bil-*?

			true?		
			no	yes	
			Trump	Bernie	att. pred.
(9)	justified?	no	# (7a)	# (7a)	non-factive know
			# (7b)	# (7b)	factive know
			✓ (7c)	✓ (7c)	think
		yes	✓ (7a)	(✓) (7a)	non-factive know
			# (7b)	✓ (7b)	factive know
			✓ (7c)	(✓) (7c)	think

Consequently, *diye bil-* can in some circumstances be translated as 'think,' (bottom three rows) but the two meanings are not strictly equivalent (top three rows). But non-factive knowledge reports retain a justification requirement from knowledge, which lacks from the expression of neutral belief.

2.2 The meaning contribution of *diye*

The previous result is interesting from an epistemological perspective. A core knowledge meaning component can be isolated, and it seems to include a justification requirement, without truth. The availability of this meaning is perhaps due to the fact that, in general, sentences of the form [p diye V] assert the existence of actual events or situations of the kind denoted by V. The examples in (10) illustrate:

(10) a. Use of *diye* with manner of speech predicates
Tunç [Bernie kazan-dı diye] {bağırdı/fısıldadı}.
Tunç Bernie win-PST.3S *diye* screamed/whispered
Tunç {screamed/whispered} that Bernie won. $\not\rightsquigarrow$ Bernie won.
(There is an actual screaming/whispering event.)

b. Use of *diye* with different belief predicates
Tunç [Bernie kazan-dı diye] {hatırlıyor/öğrendi}.
Tunç Bernie win-PST.3S *diye* remembers/learned
Tunç {remembers/learned} that Bernie won. $\not\rightsquigarrow$ Bernie won.
(There is an actual remembering/learning event.)

This seemingly trivial result rules out a potential line of analysis: It appears that *diye* does not make a factive predicate non-factive by directly operating on its meaning.

Furthermore, *diye* specifies the verbal content of the scream (whisper), in (10a), and the propositional content of a belief, in (10b). Though it looks like this content scopes under the matrix attitude predicate, evidence from a few 'inherently negative' predicates suggest scopal independence. Sentence (11a) shows a nominalization composed with the predicates 'deny' and 'falsify.' The nominalized proposition is what is denied or falsified. However, a proposition introduced by *diye*, as in (11b), denotes the content of a speech act that *accompanies* the matrix event.

(11) a. Nominalizations scope under the matrix predicate
Tunç (*bu-nu) [Bernie kazan-dığ-ı-nı] {inkar etti/yalanladı}.
Tunç this-ACC Bernie win-NMZ-3S-ACC denial did/falsified
Tunç {denied/falsified} the proposition that Bernie won.

b. *diye p* is scopally independent from the matrix predicate
Tunç (bu-nu) [Bernie kazan-dı diye] {inkar etti/yalanladı}.
Tunç this-ACC Bernie win-PST.3S *diye* denial did/falsified
Tunç {denied/falsified} this *by saying that* Bernie won.

Syntactically, it is possibile to give the matrix predicate an overt internal argument ('this') in a sentence with *diye*. This is impossibile with a nominalization.

Assuming that the facts in (11) are general, we can conclude about sentences with *diye* where factive predicates get non-factive interpretations that:

1. There is an actual event of the kind denoted by the factive predicate.
2. *Diye* introduces an independent belief predicate, and its propositional content.
3. Consequently, the factive predicate does not operate on the belief proposition.

Before moving on, note that examples like (11b), where the matrix predicate's internal argument is saturated by an overt nominal, are possible with factives too:

(12) Tunç (bu-nu) [Bernie kazan-dı diye] {biliyor/hatırlıyor/öğrendi}.
Tunç this-ACC Bernie win-PST.3S *diye* knows/remembers/learned
Tunç {knows/remembers/learned} this as Bernie winning. ↛Bernie won.

What remains to be accounted for is that the interpretation of *diye* as a speech or a belief predicate depends on the matrix predicate, and that there is a relation between the object of the matrix predicate and the belief proposition.

3 Proposal

In this section, I introduce the lexical factivity hypothesis and provide an argument for adopting it. I then sketch out a semantics for the alternants in the factivity alternation.

3.1 The lexical factivity hypothesis

A traditional way of encoding factivity in the attitude predicate's lexical entry is to say that it presupposes the truth of the propositional object it composes with [12]. This view seems to commit us to considering this propositional object as the attitude predicate's complement.The definition of a predicate like *bil-* can be written as in (13):

(13) $[\![\text{bil-}]\!]^{w}=\lambda p_{st} \,.\, \lambda x_{e} : p(w)=1 \,.\, \forall w' \; w' \in DOX(x,w) \rightarrow p(w')=1$

This function is defined only if the proposition expressed by *bil-*'s complement is true in the world of evaluation, and returns true only if that proposition is true in all of the attitude holder's belief worlds.

Under this hypothesis, factive knowledge reports with nominalizations do not pose a challenge. But, attitude reports with *bil-* are generally predicted to be factive. This is challenged by the factivity alternation. To derive non-factive reports with *bil-*, the semantics of *diye* must be such that the presupposition is suspended.

At least since Kartunnen [10], the literature acknowledges the existence of 'plugs,' which block presuppositions from projecting. One type of plug is non-factive attitude and speech predicates: If a presupposition trigger is embedded under such a predicate, the presupposition seems to fail to project. A naturally occurring example is provided in (14), where *know p* does not, to be felicitous, require *p* to be true in the world of evaluation of the sentence. (Such judgments are known to vary [10].)

(14) Sansa thinks she knows that Theon killed her two younger brothers [. . .][5]
(Theon did not kill Sansa's brothers.)

A way of accounting for the 'plugging' of the presupposition here could be to assume that the world argument of the proposition in the presupposition component is bound by the universal quantifier introduced by 'think.' Then, it suffices for the proposition expressed by the embedded clause to be true in all of Sansa's thought-worlds, which need not include the world in which the entire sentence is evaluated. This would indeed have the effect of committing the matrix subject to the truth of the proposition expressed by 'Theon killed [Sansa]'s brothers,' but not the speaker.

In Turkish, clauses (apparently) embedded under *bil-* that give rise to non-factive attitude reports are introduced by *diye*. Consequently, it would suffice to write in *diye*'s meaning whatever it is in *think* or *say*'s meaning that makes them act like plugs.In the previous section, I have provided independent evidence for this view. But, instead of 'plugging' the presupposition in the way sketched out in the last paragraph, *diye*'s semantics and mode of composition with the matrix predicate are such that the predicate does not directly operate on the belief proposition, hence being unable to trigger the presupposition of its truth. In the non-factive sentences with *diye* discussed here, the presupposition is not plugged, rather, it is not triggered.

3.2 A reason for adopting the factivity hypothesis

With nominalized clauses, the availability of a factive interpretation depends on the choice of the attitude predicate. This is illustrated by the contrasts in (15).

(15) **Context**: Trump won the election but. . .

a. # Tunç [Bernie-nin kazan-dığ-ı-nı] biliyor/öğrendi/hatırlıyor.
Tunç Bernie-GEN win-NMZ-3S-ACC knows/learned/remembers
Tunç knows/learned/remembers that Bernie won. $\rightsquigarrow$ Bernie won.

b. Tunç [Bernie-nin kazan-dığ-ı-nı] düşünüyor/varsayıyor/hayal etti.
Tunç Bernie-GEN win-NMZ-3S-ACC thinks/supposes/imagined
Tunç thinks/supposes/imagined that Bernie won. $\not\rightsquigarrow$ Bernie won.

[5] http://pickledwhale.weebly.com/blog/quite-the-little-finger-indeed

With *diye*, attitude predicates are uniformly non-factive, as shown in (16).

(16) **Context**: Trump won the election but…

a. Tunç [Bernie kazan-dı diye] biliyor/öğrendi/hatırlıyor.
Tunç Bernie won-PST.3S *diye* knows/learned/remembers
Tunç "knows/learned/remembers" that Bernie won. $\not\rightsquigarrow$ Bernie won.

b. Tunç [Bernie kazan-dı diye] düşünüyor/varsayıyor/hayal etti.
Tunç Bernie won-PST.3S *diye* thinks/supposes/imagined
Tunç thinks/supposes/imagined that Bernie won. $\not\rightsquigarrow$ Bernie won.

The difference between (15) and (16) strongly suggests that nominalizations are not inherently specified for factivity, and that whether the proposition a nominalization expresses is presupposed depends on the semantics of the embedding predicate. (For a different picture from Korean, where all nominalizations seem to be factive, in support of the external factivity hypothesis, see [18].) However, non-factivity does seem to be contributed by *diye* given that the reports in (16) are non-factive across the board.

The homophony and the external factivity hypotheses could in principle handle these facts, with the cost of making lexical stipulations and losing explanatory power.

3.3 The semantics of the alternants

The data in section 2 suggest that two pieces of meaning need to be related in attitude reports with *diye*: The event/situation introduced by the matrix attitude predicate and a secondary belief predicate introduced by *diye*, with its propositional content. I give a simplified structure associated with the sentence in (17a) in (17b), and its semantics in (17c). I assume that the attitude holder is introduced by v [13, 16], and that the knowledge (abbreviated by K) of a set of facts *causes* the propositional belief (respectively p and B), adopting the mechanism sketched out in [15] for resultatives.

(17) a. Tunç [Berni kazandı diye] biliyor.
Tunç *thinks* (lit. *knows*) that Bernie won.

b. $[_{v\text{P}}$ Tunç $[v$ $[_{\text{VP}}$ [knows it$_{12}$] [*diye* Bernie won]]]]

c. $[\![$(17a)$]\!]$=1 iff
$\exists s_1[\text{K}(\text{g}(12))(s_1)\wedge\text{holder}(\text{tunc})(s_1)]\wedge\exists s_2[\text{B}(p)(s_2)]\wedge\text{CAUSE}(s_2)(s_1)$

There is a knowledge state s_1 whose object is g(12) and whose attitude holder is Tunç, and there is a belief state s_2 whose object is the proposition that Bernie won, and s_1 causes s_2.

The internal argument of the predicate 'know' is a contextually valued nominal, which we saw could be overtly expressed in (12). Finally, it is reasonable to think that since we are dealing with connected mental states, the holder of the belief state is identified with the holder of the knowledge state introduced by v.

Now, the structure and the semantics associated with the factive attitude report in (18a) are respectively given in (18b) and (18c):

(18) a. Tunç [Berninin kazandığını] biliyor.
Tunç *knows* that Bernie won.

b. [$_{vP}$ Tunç [$_{v}$ v [$_{\text{VP}}$ knows [$_{\text{DP}}$ NMZ Bernie won]]]]

c. $[\![(18a)]\!]=1$ iff $\exists s_1[\text{K}(p)\wedge\text{holder(tunc)}(s_1)]$

There is a knowledge state s_1 whose object is the proposition that Bernie won and whose attitude holder is Tunç. (Presupposition: Bernie won.)

For simplicity, I leave out how to retrieve the propositional content of the nominalization. In (18c), the attitude predicate 'know' directly composes with the proposition that Bernie won. Given that the factive predicate is the presupposition trigger, the truth of the embedded proposition is presupposed.

This account is able to capture the meaning of *diye* used in conjunction with manner of speech predicates, if it is granted that the paraphrase in (19b) is a good approximation of the meaning of (19a).

(19) a. Tunç [Berni kazandı diye] bağırdı.
Tunç screamed *diye* Bernie won.

b. Tunç's screaming caused him to say (*diye*) that Bernie won.

Data from the previous section motivate the need for an additional interpretation of *diye* as introducing content of a speech act. I must leave a formal implementation of this variability (illustrated also in (20)) for further research.

I would like to close this section by bringing additional plausibility of a causal interpretation for *diye*. Turkish has uses of *diye* other than in attitude reports. It introduces a reason, in (20a), or a purpose clause, in (20b).

(20) a. Tunç [Berni kazan-dı diye] ağladı.
Tunç Bernie win-PST.3S *diye* cried
Tunç cried because (*diye*) Bernie won.

b. Tunç [Bernie kazan-sın diye] ağladı.
Tunç Bernie win-OPT.3S *diye* cried
Tunç cried so that (*diye*) Bernie would win.

Many details remain to be worked out, but the present account paves the way for a unified treatment of *diye*.

3.4 The cross-linguistic perspective

Catalan, Greek, Hungarian [1] and Korean [18], are languages that are reported to display a factivity alternation like in Turkish. Further research is required to see whether some of these languages pattern like Turkish, and whether the present analysis could be extended to them.

From a cross-linguistic perspective, the meaning associated with *diye bil-* seems related to ones described by Kierstead [11], for Tagalog *akala*, and Glass [6], for

Mandarin *yǐwéi*. *Akala* and *yǐwéi* are belief predicates reported to express mistaken belief when their attitude holder is a third person. If the attitude holder is also the speaker, Glass reports that the result is a *hedgy* belief report, rather than a *mistaken* one. I do not discuss first person attitude holders here in the interest of space, but Turkish is similar to Mandarin in that non-factive 'know' in the first person signals that: the speaker$_i$ is justified in believing that *p* but that they$_i$ are open to the possibility that *not p*. Furthermore, data from a native speaker consultant (Hsin-Lun Huang, p.c.) suggests that justification plays a crucial role in licensing *yǐwéi*, which I showed is also the case for *diye bil-*.

The Turkish facts described here are an interesting addition to the Tagalog and Mandarin data, given that although *diye bil-* is sometimes used to express mistaken belief, this is not always the case. Illustrated in (21) is a use of *diye bil-* where the speaker lacks knowledge about *p*, but asserts that a third person is justified in believing that *p*, whereby presenting evidence *in favor of p*. This kind of use is consistent with the 'justified but not necessarily true' description of *diye bil-*, and it suggests that Glass's account of *yǐwéi* does not straightforwardly extend to Turkish.

(21) **Context**: The speaker is asked: "Who won the election?"
Valla ben bilmiyorum ama Tunç Bernie kazandı diye biliyo.
tbh I don't.know but Tunç Bernie won *diye* knows
To be honest, I don't know, but Tunç *thinks* (lit. *knows*) Bernie won.

Consequently, *diye bil-* does not mean *falsely believe* (though this is sometimes an attested inference). The similarities observed at the onset of this subsection could be an effect of *diye* having the semantics of 'believe,' which is known to give rise to falsity inferences, in appropriate contexts, cross-linguistically [19].

Finally, the alternation in the truth requirement is observed for an attitude predicate otherwise used to express *knowledge* (compare Tagalog *akala*, '(falsely) believe' to *alam*, 'know,' and Mandarin *yǐweí* to *zhīdaò*) and productively extends to other attitude predicates. Non-factive uses of otherwise factive predicates seem to depend on the syntax of the embedded clause, and on the semantics of *diye*, rather than on lexical idiosyncracies associated with the predicates themselves.

4 The two alternative hypotheses about the factivity alternation

In this section, I discuss two competitors of the lexical factivity hypothesis argued for in section 3.

4.1 The homophony hypothesis

The homophony hypothesis is that the Turkish lexicon contains two homophonous attitude predicates *bil*-$_{\text{FACTIVE}}$ and *bil*-$_{\text{NON-FACTIVE}}$.

The two predicates select for different kinds of propositional objects, respectively a nominalized and a tensed clause, that have in common the feature of forming the content of a justified belief. The predicate *bil*-$_{\text{FACTIVE}}$ further imposes the condition that the object denote a true proposition, whereas *bil*-$_{\text{NON-FACTIVE}}$ comes with no such condition. The former corresponds to the familiar factive predicate 'know.' The latter is a more unusual predicate, yet one that fits cross-linguistic patterns.

Hsin-Lun Huang (p.c.) reports that *yǐwéi*, like *diye bil-*, requires the reported belief to be justified (in addition to implying that the belief is false, in Mandarin, as described by Glass [6]). The existence of two distinct attitude verbs, one for knowledge (*zhīdaò*) and the other for justified (yet perhaps false) belief, could be seen as providing support for the homophony hypothesis. Turkish would have both 'know' and a predicate like *yǐwéi*, which happen to be pronounced the same. A similar reasoning could apply to all predicates participating in the alternation: *hatırla*-$_{\text{NON-FACTIVE}}$ and *hatırla*-$_{\text{FACTIVE}}$ for 'remember,' etc.

I do not see an easy way of dismissing this hypothesis on empirical grounds. In the absence of a strong motivation in favor of it, this is perhaps enough to set it aside.

4.2 The external factivity hypothesis

The final hypothesis, 'external factivity,' states that *bil-* does not encode factivity in its lexical entry at all. (Keeping to the particular construal of factivity discussed up to now, this can be reformulated as: *bil-* is not a semantic presupposition trigger.) Such an account is argued for by Hazlett [7, 8].

The first question that this hypothesis raises is what the meaning of *bil-* is. In the data section, we observed that justification sets belief reports apart from knowledge reports. This fact could be written down in the lexicon, informally as in (22):

(22) S *bil-* p := S has the justified belief that p
for S, an attitude holder, and p, a proposition

The definition encodes the justification requirement, and remains silent about the truth of p. This gives us a way of accounting for the non-factive uses of *bil-*, uses that are distinct from plain belief or thought reports. But, a consequence of the external factivity hypothesis is that we must have a way of generating factivity compositionally. This is a requirement given that *bil-* gives rise to factive readings with nominalized clauses. The question then is what, in the semantics, has the potential to introduce factivity? Drawing on work by Kratzer [16, 17] and others [2, 18], the external factivity hypothesis receives a particular implementation with the assumptions that: attitude predicates do not take clauses directly as their syntactic complement, and that modality is introduced from within the embedded clause, by the complementizer. This is appealing, but the facts discussed here do not, I believe, provide sufficient evidence for this view.

5 Concluding remarks

This paper introduces novel data from Turkish attitude reports that suggest that certain predicates including *bil-*, 'know,' are interpreted as factive or as non-factive depending on the syntactic type of the propositional object they combine with. I have argued that factivity is ‘concealed’ by the functional element *diye*, which introduces an independent belief predicate along with its propositional content. Given that beliefs are not factive, neither are attitude reports that make use of *diye*. In the absence of *diye*, factive presuppositions are triggered as usual, by factive attitude predicates directly taking scope over the (nominalized) proposition they embed.

References

1. Abrusán, M.: Predicting the presuppositions of soft triggers. Linguistics and Philosophy 34(6), 491–535 (2012)
2. Bogal-Allbritten, E.: Building meaning in Navajo. Ph.D. thesis, University of Massachusetts, Amherst (2016)
3. Şener, S.: Non-Canonical Case Licensing is Canonical: Accusative Subjects of CPs in Turkish (2008), ms. University of Connecticut
4. George, L., Kornfilt, J.: Finiteness and Boundedness in Turkish. In: Henry, F. (ed.) Binding and filtering, pp. 105–127. MIT Press (1981)
5. Gettier, E.L.: Is justified true belief knowledge? Analysis 23(6), 121–123 (1963)
6. Glass, L.: The negatively biased Mandarin belief verb *yiwei* (2016), http://ling.auf.net/lingbuzz/002600/
7. Hazlett, A.: The myth of factive verbs. Philosophy and Phenomenological Research 80(3), 497–522 (2010)
8. Hazlett, A.: Factive presupposition and the truth condition on knowledge. Acta Analytica 27(4), 461–478 (2012)
9. Karttunen, L.: Some observations on factivity. Papers in linguistics 4(1), 55–69 (1971)
10. Karttunen, L.: Presuppositions of compound sentences. Linguistic Inquiry 4(2), 169–193 (1973)
11. Kierstead, G.: Shifted indexicals and conventional implicature: Tagalog *akala* ‘falsely believe’ (2013), talk presented at SALT 23, UCSC
12. Kiparsky, P., Kiparsky, C.: Fact. In: Bierwisch, M., Heidolph, K.E. (eds.) Progress in Linguistics. The Hague: Mouton (1970)
13. Kratzer, A.: Severing the external argument from its verb. In: Phrase structure and the lexicon, pp. 109–137. Springer (1996)
14. Kratzer, A.: Facts: Particulars or information units? Linguistics and philosophy 25(5), 655–670 (2002)
15. Kratzer, A.: Building resultatives. Event arguments: Foundations and applications pp. 177–212 (2005)
16. Kratzer, A.: Decomposing attitude verbs. Talk given at The Hebrew University of Jerusalem (2006)
17. Kratzer, A.: Modality for the 21st century. In: 19th International Congress of Linguists. pp. 181–201 (2013)
18. Moulton, K.: Natural selection and the syntax of clausal complementation. Open Access Dissertations p. 99 (2009)
19. Percus, O.: Antipresuppositions. In: Ueyama, A. (ed.) Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science, vol. 52, p. 73. Kyushu University (2006)
20. Simons, M.: On the conversational basis of some presuppositions. In: Perspectives on linguistic pragmatics, pp. 329–348. Springer (2013)
21. Simons, M., Beaver, D., Roberts, C., Tonhauser, J.: The best question: explaining the projection behavior of factives. Discourse Processes (2015)

Be-gadol (~basically) as a question sensitive operator

Moria Ronen
Bar Ilan University
Mori.me@gmail.com

Abstract. The notion of "good answer" has been discussed in length in recent years. The present paper addresses a Hebrew hedger, *be-gadol*, roughly translated as *basically*. It suggests that *be-gadol* is a lexical item whose distinguishing function is to convey a restriction on the context of utterance concerning a relation between answers to the QUD on an answerhood scale, built using notions such as informativity and resolution (Roberts 1996; Ginzburg 1995; van Rooij 2003). This significantly supports the linguistic reality of these notions.

Keywords: Hedging, decision problems, resolution, questions, answers.

This paper deals with the Hebrew modifier *be-gadol*. We propose that *be-gadol* combines with a proposition *p*, and conveys that this *p* is not a good enough answer to the question under discussion (QUD), and that there is another answer, p_{best} which is better than *p* and close to it on an answerhood scale. To measure the "goodness" of an answer we use scales based on theoretical tools like informativity (following Roberts 1996) and resolution (following Ginzburg 1995, van Rooij 2003) and suggest that *p* is less informative and also less helpful than p_{best}, as detailed below.

The structure of this piece is as follows: the first section brings the main uses found with *be-gadol* as well as some central observations; section 2 discusses two relevant types of theories, one dealing with the notion of QUD, and another regarding the concepts of resolution and decision problems. Section 3 presents the suggestion regarding *be-gadol*, and application to two types of its uses. Section 4 brings forth an apparent problem and a QUD-based solution, and section 5 concludes.

1 Data

Be-gadol (literally 'in big') is a Hebrew hedger which has a variety of uses. It is almost never unacceptable. In this paper we focus on its occurrences in declaratives[1]. Consider for example the main uses in (1) and the corresponding interpretations in (3a-d). Modifying (1) slightly to future tense (2) yields yet another use (3e)[2]:

[1] *Be-gadol* can also occur in questions, but rarely in imperatives. These cases merit future research.

[2] For space reasons only the answers are given in Hebrew

1. A: What did Rina do in the party?
   B: Hi be-gadol rakda
   She in.big danced
   'She ~basically danced.'
2. A: What will Rina do at the party?
   B: Hi be-gadol tirkod
   She in.big will.dance
   'She will ~basically dance.'
3. Interpretations of (1):
   (a) The temporal use: She danced most of the time.
   (b) The significance use: The most important thing she did was dance (crucially, even if this did not occupy most of the time).
   (c) The not-enough-details use: She danced a specific kind of dance, e.g. tango, which we don't specify.
   (d) The approximative use: She danced in a non typical way, e.g. she swayed gently from side to side in a way which might still count as dancing.

An additional interpretation of (2):

(e) The reduced commitment use: she promised she will dance / it seems that she will dance, but don't count on it[3].

To better demonstrate the not-enough-details use, consider also (4), where Rina actually does something more specific than to be an alternative energy engineer, e.g. she deals with extracting fuel from corn:

4. A: What do you do?
   B: Be-gadol ani mehandes energiya xalufit
   In.big I engineer energy alternative
   '~Basically I'm an alternative energy engineer.'

Importantly, the different readings are independent, e.g., (3b) can be true when (3a) is false and vice versa. While some of the uses can be paraphrased with other particles (e.g. *for the most part*, *more or less*, or *probably*), the semantics of *be-gadol* cannot be reduced to that of any of these particles, since none of them captures all the uses. For example, (5a) can paraphrase (3a) but not (3b-e), (5b) can paraphrase (3d) but not (3a-c, e), and (5c) can paraphrase (3e) but not (3a-d).

5. Possible paraphrases

   (a) For the most part, she danced.
   (b) She more or less danced.

[3] Also consider (i) where B conveys that she intends to come, but cannot guarantee it.
(i) A: are you coming to the party tomorrow?
B: Be-gadol yes

(c) She will probably dance.

Crucially, the felicity of *be-gadol* depends on context in a specific way. Consider for example:

6. Sarah: What did Rina do at the party?'
   (a) Miri: Hi rakda
   She danced
   She danced
   (b) Miri: Hi rakda ve.Sara
   She danced and.sang
   'She danced and sang.'
   (c) Miri: Hi rakda, Sara ve.halxa la.sherutim
   She danced, sang and.went to.the bathroom
   'She danced and sang and went to the bathroom.'

Given a context of a casual discussion between friends, *be-gadol* seems to be felicitous with (6a), but not with (6b) (assuming that (6b) or even (6c) exhausts Sarah's actions in the party). In contrast, given a different context, e.g., where Miri is a detective reporting to her employer Sarah, *be-gadol* is better with (6b) but not with (6a,c) (assuming that (6c) exhausts Sarah's actions in the party). *Be-gadol*, then, seems to be sensitive to the roles and goals of participants in the discourse. These factors have been identified as crucial for the notion of resolving questions. This will become crucial for the analysis in section 3.

Hence, notwithstanding the variety of uses and due to the specific context sensitivity observed, we propose the following: *be-gadol* denotes a single operation in all its uses. In particular, in all the uses the presence of *be-gadol* indicates that p is not the best answer to the QUD. The 'best answer', p_{best}, is better than the prejacent p since it is more informative than p and more 'helpful' (as will be explained shortly). In addition, p_{best} cancels a default implication of p. Together, these components lead to the flexible hedging effect of *be-gadol*.

To illustrate, consider the use of *be-gadol* in (1). We assume that the prejacent of *be-gadol*, i.e. (6a), has a default implication q as in (7a), although it is possible to cancel this implication and universally quantify over the times in a narrower domain. We refer to the interpretation without the implication as p, as in (7b). Thus, (6a) denotes p but only defeasibly implies q. By contrast, p_{best} is any proposition entailing p and *not* q, e.g. as in (7c), where it is also further specified what Rina did when she didn't dance (e.g., sing). Thus, p_{best} is more informative than p because p leaves open the question of whether Rina danced in all the times in the default domain or only in a subset of it. In contrast, p_{best} entails that she only danced in the narrower domain. Being more informative, p_{best} is a more helpful answer to the QUD in (6a) than p. At the same time, p_{best} remains implicit, and this fact together with the cancelation of the strong default prejacent implication (7a) in favor of the weaker prejacent interpretation (7b) creates the hedging effect.

7. (a) q = Rina danced in all the times in some default domain (cf. von Fintel 1994).
   (b) p = Rina danced in all the times in the default domain or only in a narrower (further restricted) domain.
   (c) $p_{best} \Rightarrow p \wedge \neg q$ = Rina danced in all the times in the narrow domain, and not in a default domain.

Notice that using informativity alone is not sufficient. To demonstrate this, assume that *p* is less good than p_{best} because p_{best} is just more informative than *p*, i.e., only asymmetrically entails *p*. For example, in the context of question (6), let p_1 be (6a), p_2 be *Hi halxa la.sherutim* ('She went to the bathroom'), and p_{best} be (6c). The best answer p_{best} entails both p_1 and p_2. However, in a casual conversation *be-gadol* is only felicitous with p_1 as its prejacent (though when reported by a detetctive *be-gadol* is also felicitous with p_2). Intuitively, p_2 is not relevant to the speakers' goals in ordinary contexts, and therefore is not licensed as a prejacent of *be-gadol*. Thus, the role of contextual goals has to be represented. To develop an account along these lines, we now move to the theoretical tools.

2 Theoretical tools

According to Roberts (1996), once a question was explicitly or implicitly raised in discourse, it becomes a member of the set of questions under discussion (QUDs), which the participants always try to answer. Roberts relies on the notion of questions as denoting sets of answers (Hamblin 1973) and says that a complete answer enables the evaluation of the truth value of each proposition in the set of potential answers. A partial answer enables the evaluation of at least one proposition in the set. For example, the question *What did Rina eat?* denotes a set of answers, e.g. {*Rina ate a tomato, Rina ate pizza, Rina ate a sandwich…*}. In a scenario where Rina ate pizza and a tomato, a partial answer would be *Rina ate pizza*, and a complete answer can be *Rina ate pizza, and a tomato, and she didn't eat a sandwich.* In this sense, a complete answer is 'more informative' than a partial one.

Another branch of theories dealing with answers to questions have to do with the notion of 'resolution', making use of theories regarding decision problems. In particular, Ginzburg (1995) shows that a good 'resolving' answer is an answer which takes into consideration the roles and goals of the interlocutors.

To formalize this notion of resolution, van Rooij (2003) uses decision theory. Intuitively, speakers use discourse to resolve **decision problems**, i.e., to decide what to do to achieve their goals. Decision problems have three components. The first is the **beliefs** the agent has regarding this world (i.e. what is less / more likely); the second is the different **actions** available for the agent, and the third is her **preferences** based on the first two factors, i.e., her preferred outcomes among the outcomes of all possible actions given how she views the world. The preferences factor encodes the goals of the agent (what she aims to achieve in the world), given her possible actions and beliefs. These three components of a decision problem are used to calculate the 'expected utility' of actions, and consequently the helpfulness of different pieces of information (e.g., propositions expressed as answers to questions) in resolving the decision problem (the problem of rationally deciding between different actions), as detailed below. **Utility** is used to capture the helpfulness of a piece of information, in other words how much this piece of information narrows down the set of possible actions $A= \{a_1, a_2, a_3,...a_n\}$ from which the agent has to choose, as follows. Utility is calculated using a function from actions and worlds to real numbers. Each action which is a member of the set *A* has a utility value in a world *w*, namely, *U(a,w)*. For example:

8. Rina is throwing her son a party. The children are all over the age of five, in a world where children under five only like balloons, and children over five only like clowns. The utility function reflects the desires and wishes of the agent:

   (a) Given that the agent wants the children to be happy, the utility of getting a clown for the party > the utility of ordering balloons.
   (b) Given that the agent doesn't want the children to be happy, the utility of getting a clown for the party < the utility of ordering balloons.

However, in reality the agent doesn't know what the actual world is like. Had she known, it would have been clear which action she should choose. This uncertainty is realized by the function *P* (probability function). Let W be the set of worlds. Worlds that are still candidates to be the actual world have probability higher than zero. Importantly, the probability of all these worlds has to add up to 1. Van Rooij uses these measures to model a decision problem as a tuple *<P,U,A>* where *P* stands for probability, *U* for utility and *A* for a set of actions. The *expected utility* (EU) of an action *a* is thus defined as the sum of a's utilities in all worlds weighed by their probability:

9. $EU(a) = \Sigma_w P(w) \times U(a,w)$

The utility of the action with the maximal expected utility, i.e. the one which is most likely to be most helpful, is as follows:

10. $maxEU(P,U,A) = Max(\{\Sigma_w P(w) \times U(a,w): a \in A\})$

The rational choice is of the action with this value, namely that action a for which *EU(a) = maxEU(P,U,A)*. However, sometimes several actions may have this value, and thus more information is needed to make a decision. Van Rooij then uses the EU of different actions to measure and compare the utility of different pieces of information, to learn which one is more helpful in order to choose an action. To receive more information the agent asks a question, and is given an answer. Following that, she can update the probability function over worlds, recalculate the expected utilities, and re-determine the action with the maximal EU. Let $P_C(w)$ be the conditional probability of the world *w* given answer C (see van Rooij 2003: page 9). The value of the action with the highest EU given C is then as follows:

11. $maxEU(P_C,U,A) = Max(\{\Sigma_w Pc(w) \times U(a,w): a \in A\})$

To wrap up, the EU of the action with the maximal EU given C is generalized summing over its utility in each world, times the probability of this world being the real world given C. Learning C can change the probability distribution. A proposition C is thus 'resolving' iff it leaves the agent with exactly one action with a maximal expected utility, namely one action which is the best one to choose. Consider for example (12) given the two pieces of discourse in (13a,b):

12. Rina wants to know whether to take an umbrella if she goes for a walk outside. Possible actions:
   (a) To take an umbrella.
   (b) Not to take an umbrella
13. What does it look like outside?
   (a) It's sunny
   (b) It's cloudy

Upon hearing C in (13a) Rina knows that the probability that it is raining is lowered, as is the expected utility of action (12a) (since worlds in which the utility of taking an umbrella is high become less probable). In contrast, given (13b) the situation is reversed (since worlds in which the utility of taking an umbrella is high become more probable). Based on these definitions, van Rooij defines several possible ways for a proposition C to be better than another proposition D[4]. We adopt his definition with slight differences to capture the meaning of *be-gadol*. Let A* be the set of propositions of the form "One should choose action a" for each action a in A. Let C_{A*} be the subset of propositions p in A* which are consistent with C ($C \cap p = \varnothing$). C is better than D, $C >_{dp} D$ (where dp stands for a decision problem) iff one of the following conditions holds:

14. (a) C leaves the agent with fewer actions than D does: $|C_{A*}| < |D_{A*}|$. Or:
   (b) The maximal expected utility is higher given C than it is given D: $maxEU(P_C,U,A) > maxEU(P_D,U,A)$. Or:
   (c) D asymmetrically entails C, but it is not more helpful than C in any way: $|C_{A*}|=|D_{A*}|$ and $maxEU(P_C,U,A)= maxEU(P_D,U,A)$. Thus, D is less good because it is over-informative.

We propose that if agents need to rank two proposition according to their goodness as answers to a question (relative to a decision problem), they use these conditions successively. They first check whether (14a) is met. If not, they turn to check whether (14b) is met, and if it isn't they turn to (14c).Finally, we say that the goodness of a proposition D is close to that of a better proposition C, $Close>_{dp}(C,D)$, iff D is almost as good as C in one of the following senses:

[4] Anton Benz (2007) has argued that we actually need to take a speaker into account, i.e., that bare-naked decision problems are not enough, in order to define "goodness" of answers (see also Franke & de Jager (2012) for additional diuscussion).

15. (a) Small ($|C_{A*}|$ - $|D_{A*}|$) Or:
(b) Small (maxEU(P_C,U,A) - maxEU(P_D,U,A)).

With these tools we are now in a position to propose an account of *be-gadol.*

3 Proposed account of *be-gadol*

As indicated above, we intuitively propose that *be-gadol* combines with a proposition *p* and indicates that *p* is not the best answer to the QUD but it is still a good answer, i.e. close to the best answer on an answerhood scale. Also, *p* implies a proposition *q*, while p_{best} entails $\neg q$. In attempt to make this intuition more precise, we propose the lexical entry in (16), which will be followed by a short discussion of the status of its components:

16. $[[\text{Be-gadol}]]_{w,dp}= \lambda p \in \text{QUD}.\ \lambda w: \exists q\ [q \in \text{QUD} \wedge (p \sim\sim> q) \wedge \text{BEST} \neq \varnothing$, where
$\text{BEST} = \{p_{best} \in \text{QUD}: (p_{best} \Rightarrow p \wedge \neg q) \wedge \text{resolving}_{dp}(p_{best}) \wedge \text{best}_{dp}(p_{best})$[5]
$\wedge\ (p_{best} >_{dp} p) \wedge \text{Close}_{dp}(p_{best}, p)\}].\ \exists p_{best} \in \text{BEST}\ [w \in p_{best}]$.

In words, *be-gadol,* which is interpreted with respect to a decision problem and a world, takes a proposition *p* which is a member of the QUD and a world *w* and is defined iff there is another proposition *q*, also a member of the same QUD (i.e. it is an answer to the same question), which is implied, but not strictly entailed, by *p*.[6] In addition, there is a non empty set of propositions, also answers to the QUD, BEST. Each proposition p_{best} in BEST entails both *p* and $\neg q$, and with respect to the decision problem each one is a resolving answer, the best answer, better than *p* on a resolution scale, but still close to *p*. If defined, *be-gadol p* is true in *w* iff there is an answer p_{best} in BEST which is true in *w*.[7]

With regards to the status of the different components, we ran the family of sentences tests for the components in (16)[8]. To illustrate, consider (17):

17. (a) ha-xeder be-gadol naki (the room is ~basically clean)
(b) p= The room is mostly clean.

[5] There might be no resolving propositions in some decision problems. Consider an arbitrary decision problem and then add, for each action, another action with the same utilities. No proposition will single out a single act as best. This account predicts that *be-gadol* would not be felicitous in such cases. Otherwise, we will have to revise the definitions

[6] q is a default interpretation of p; one way to formalize this is by stating that most p worlds are q worlds.

[7] We still need to check the hypothesis that $\neg q$ is only an implicature of *be-gadol p*, that is: *(p_{best} ~~> $\neg q$) $\wedge$ ($p_{best} \Rightarrow p$)* (or alternatively *($p_{best} \Rightarrow p \wedge \neg q$)* and *be-gadol p* asserts that *p* and implicates that $\neg q$).

[8] It is difficult to embed *be-gadol* under negation, so *the room isn't be-gadol tidy* is marginal for many speakers. This might result from the PPI characteristics approximators sometimes show, e.g. *almost* (Partee 2004). As a result we checked the components only under conditionals and questions.

(c)$\neg$q= The room is not completely clean
(d) p~~>q='The room is clean' defeasibly implies but not entails 'The room is completely clean'
(e) p_{best}[9]= The room is clean but not completely clean
(f) p_{best} is resolving
(g) $_{pbest}$ is the best answer to the QUD
(h) $p_{best} >_{dp} p$
(i) $Close_{dp}(p_{best}, p)$
(j) $p_{best} \Rightarrow p\wedge\neg q$
(k) p_{best} is true w.

For each component in (17b-k) we checked whether it is implied from the basic form of the sentence as in (17a) versus the embedded form of the sentence, as in e.g. *Is the room be-gadol clean?*. The results suggest that (17b,c,k) are asserted, i.e. these three components are implied by the basic form, but don't survive under embedding. The rest of the components seem to be different and to hold (or not) independently of the truth or falsity of (17a) and its embedded correspondents.

Notice that the semantics of *be-gadol* depends on the availability of a better answer p_{best}, which can be realized as a more informative answer. A more 'informative' answer means, intuitively, that the prejacent *p* of *be-gadol* leaves something out (it is underspecified). To illustrate, the application of the proposal in (16) to cases like (1)-(2) is as follows. Quantification is restricted to some contextually salient domain, which can be called $D_{default}$. For the temporal reading in (3a) (she danced most of the time of the party) the use of *be-gadol* implies that Rina danced in all the times in some domain D_{narrow}, which is a proper subset of this domain $D_{default}$ (e.g., the entire time of the party). In addition, there is a time in the difference $D_{default} - D_{narrow}$ in which she did something else which can be left unspecified, as stated in (18a). The fuller and better answer p_{best} is the proposition expressed by a more informative clause, such as (18b). Importantly, as *p* and p_{best} have to be close, i.e. *p* has to still be a good answer to be a felicitous prejacent of *be-gadol*, in this case the narrower domain has to be large enough to consist of most of the time Rina attended the party.

18. Sarah: What did Rina do at the party?
Miri: She be-gadol danced
(a) p = Rina danced in all the times in domain D_{narrow}. In a superset of that domain, $D_{default}$, she either danced or sang.
(b) q = Rina danced in all the times in a superset of that domain, $D_{default}$, which encompasses more party time.
(c) p_{best}= In all the times in D_{narrow} Rina danced. There is a superset, $D_{default} - D_{narrow}$, in which she didn't dance.

To check the conditions in (14) for examples like (18), consider a case where Sarah is trying to decide whether to invite Rina to join a dancing group. We can have two

[9] This is one possible example of p_{best}. Another is, e.g. 'The room is clean except for the windows'

scenarios, one where Sarah only invites to the dancing group people who dance all of the time of the party, and another where she invites also people who dance most of the time of the party. In the first scenario, only (18c) is resolving and condition (14a) holds and p_{best} is better than p since only p_{best} is resolving. In the second scenario, both (18a) and (18c) are resolving, since hearing both would result in inviting Rina. In such a scenario condition one (14a) is not fulfilled, but given (18a) Sarah would be less sure that Rina would enjoy the dancing group, hence the EU of inviting her given (18c) would be higher than given (18a). Thus, the second condition (14b) is fulfilled and p_{best} would be better than p. In both scenarios the cancelation of the q implication would have an effect, in the first it would stop p from being resolving, and in the second it would lower the EU of inviting Rina.

Temporal interpretations like the one in (18) are **quantificational**[10]. In contrast, the uses (3b-d) are **not quantificational**. Consider (19), with examples for *q, p* and p_{best} of the non-quantificational reading (5). (19d) is an example of an overinformative answer.

19. Miri: what do you do?
    Rina: Be-gadol I'm an alternative energy engineer
    (a) p=Rina is some an alternative energy engineer. She might be a certain type of an alternative energy engineer, but doesn't have to be.
    (b) q= Rina is a general sort of an alternative engineer.
    (c) p_{best}= Rina is a specific kind of an alternative engineer (e.g. she deals with extracting fuel from corn.)
    (d) $p_{overinformative}$ = Rina is a specific kind of engineer (e.g. she deals with extracting fuel from corn), and she paints for a hobby.

If Sarah is a manager looking for an engineer for her solar energy department, (19a) would not be enough to tell her whether she should hire Rina, while (19c) would.[11] p_{best} is more helpful than p with respect to this decision problem as it leaves the agent with fewer actions. Importantly, p_{best} is not too informative[12]. For example, (19d) cannot function as p_{best}, although (a) like (19c) it is resolving, and (b) (19d) has the same maximal expected utility with respect to this decision problem as (19c). That is, in the indicated context (19d) and (19c) are equally helpful. However, (19d) asymmetrically entails (19c), and therefore according to (14c), it is less good. This in turn means that

[10] Other uses like (3e) can be also characterized as quantificational. The interpretational differences arise from differences in the entities quantified over; e.g., the future reading in (3e) can be taken to involve quantification over worlds. Generally, in quantificational cases like (3a), where the prejacent p is of the form Det($A \cap D$,B), and D is some contextually determined domain, *be-gadol p* implies that for some salient domain $D_{default}$ and its proper subset D_{narrow}, Det($A \cap D_{narrow}$,B) is true, and Det($A \cap D_{default}$,B) is false.

[11] The 'closeness' condition in such cases still has to be explored.

[12] Notice that stronger statements are not necessarily more resolving even if they are not over informative but rather underinformative (see Franke & van Rooij 2015, section 2.1).

it is not best relative to $>_{dp}$ as the semantics of *be-gadol* requires (because ∃p' s.t. $p' >_{dp} p_{best}$). In sum, the condition in (14c) assures that p_{best} is not more informative than necessary to resolve the decision problem.

Although domain restriction is not present for non quantificational uses, the notion of 'some implication *q* relevant for the decision problem which is false' seems to serve as a common denominator. For example, in the "not enough details use" in (19a) (she is some kind of an alternative energy engineer), it regards the unspecified identity of the exact nature of her work. For the temporal use in (18a) it regards the exact nature of her actions in the difference between D_{narrow} and $D_{default}$.

4 An apparent problem and a QUD-based solution

In addition to the uses in (1-2) and (4), we find another use which seems to challenge our present analysis. This use, which we call the 'change-your-question' use, is illustrated in (20) and in Greenberg & Ronen (2013):

20. Context: Dani is a mutual friend of Miri and Rina. He tried to get into medical school, which involves a test and an interview. If you pass the test, you most likely pass the interview.
(a) Miri: Did Dani pass the test?
(b) Rina: Be-gadol hu avar, aval hu nixSal ba.reayon
In-big he passed but he failed the.interview
'Be-gadol he passed, but he failed the interview'[13].

Given our analysis, *p = Dani passed the test* should imply some *q* which is negated by p_{best}. However, the identity of *q* and p_{best} is unclear given the specific question in (20a)., so no p_{best} better than *p* seems available. Despite this, independently motivated mechanisms discussed in the literature on questions serve to account for the felicity *be-gadol* in such cases as well. These mechanisms are discussed in relation to projective meanings and the *at issue/not at issue* distinction. In particular, Simons et al (2010) suggest that usually projection of presupposition is blocked when the content is at-issue relative to the QUD. However, they bring some examples where such projection is possible. Consider (21) below, a slight variation of the original example in Simons et all, which is odd without its context.

21. Context: Chloe is writing invitations to her birthday party to kids in her class. Her mother notices that all of the invitations are to girls.
(a) Mother: Are there any boys in your class?
(b) Chloe: I don't like the boys in my class.

That there are boys in Chloe's class is the answer to the question in (21a), thus at issue content. This content is nonetheless projected in (21b). However, Simons et al. (2010) claim that this is not a counter example for their generalization that content

[13] Whether the presence of the 'but' phrase is necessary or not is an issue we will explore in the future. The prejacent for the sake of this piece of work is without it.

does not project when at-issue. They say that the answer in examples like (21b) addresses not only the immediate QUD, but also a broader QUD. In (21b) Chloe does not only answer the direct question which is explicitly asked, but also an implicit question along the lines of "Why aren't you inviting any of the boys to your party?". This accommodation of an additional implicit question based on the speakers' contextual goals makes the sequence in (21) felicitous since the non-projective content answers the broader question. We propose that the same mechanism of QUD shifts can be applied to solve our problem as well. Consider Rina's utterance in (20) interpreted not with respect to Miri's narrower question, but with respect to a broader question (22a) as follows:

22. Miri: Did Dani pass the test?
 (a) Broader implicit question: Did he get accepted?
 Rina: be-gadol he passed, but he failed the interview.
 (b) p= Dani passed, he most likely got accepted, but maybe not.
 (c) q= Dani got accepted
 (d) p_{best}= Dani passed but didn't get accepted.

In the context of the broader implicit QUD, if Miri wants to decide whether she should congratulate Dani or not on his acceptance to medical school, (22b) would not resolve this decision problem in the best way. The best resolving answer (22d) negates the strong implication (22c) leaving Miri with a resolved decision problem of not congratulating Dani[14].

5 Summary

To sum, we report on a particle which conveys that its prejacent is a good but not best answer to the QUD. The best answer p_{best} is more informative and more resolving than p. The very identification of a lexical item whose basic semantics resorts to the notion of resolution is novel. This particle also demonstrates shifts to broader questions, which has been identified independently in other constructions. The understanding of such particles can thus enrich the general understanding of hedging and answerhood.

In the future, this analysis needs to be refined and additional aspects have to be addressed. For example, the status of different components e.g. the (entailed or implied) status of $\neg q$ by p_{best} requires further clarification, as does the nature of the association between *be-gadol* and focus. An additional issue is why the 'reduced commitment' use is restricted to future tense, and why cooperative speakers would use *be-gadol* to begin with. Last but not least, it seems that this analysis can be further extended to a family of similar hedges in various languages (e.g., *basically*, *ekronit, be-ikaron*). An emerging question is whether there exists a family of 'answerhood – sensitive' particles, which parameters they share, and which distinguish between them.

[14] The 'closeness' condition in this cases still has to be explored as well.

6 References

1. Benz, A., 2007. On relevance scale approaches. In: Puig-Waldmuller, E. (Ed.), Proceed- ¨ ings of Sinn und Bedeutung 11. Universitat Pompeu Fabra, Barcelona, pp. 91–105.
2. von Fintel, K., (1994) Restrictions on Quantifier Domains. Ph.D. dissertation, University of Massachusetts at Amherst
3. Michael Franke and Robert van Rooij (2015). In: Models of Strategic Reasoning: Logics, Games and Communities, Ed. by Johan van Benthem, Sujata Gosh and Rineke Verbrugge, Heidelberg, Springer, Chapter 8
4. Michael Franke, Tikitu de Jager, and Robert van Rooij (2012). "Relevance in Cooperation and Conflict". In: Journal of Logic and Computation 22.1, pp. 23-54
5. Ginzburg, J., (1995) Resolving Questions, Parts I and II. Linguistics and Philosophy 18.5: 459-527, 18.6:567-609.
6. Greenberg, Y., and Ronen, M. (2013). Three approximators which are almost/more or less/be-gadol the same. Proceedings of IATL28.
7. Nakanishi, K., and Romero. M., (2004). Two constructions with Most and their semantic properties. In Proceedings of NELS 34, ed. Keir Moulton and Matthew Wolf, 453–467. Amherst, Mass: GLSA
8. Partee, B. H. (ed) (2004) The Airport Squib: Any, Almost, and Superlatives, in Compositionality in Formal Semantics, Blackwell Publishing Ltd, Malden, MA, USA.
9. van Rooij, R., (2003). Questioning to resolve decision problems. Linguistics and Philosophy: 727–763.
10. Roberts, C., (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In: Jae Hak Yoon — Andreas Kathol (eds): OSU Working Papers in Linguistics 49: Papers in Semantics, 91–136. The Ohio State University, Columbus.
11. Ronen, M., (2014). It's a big word after all-, the Hebrew hedger be-gadol. Master's thesis. BIU.
12. Simons, Mandy, Judith Tonhauser, David Beaver, and Craige Roberts. (2010). What projects and why. In Semantics and Linguistic Theory (SALT) 21, 309–327. Ithaca, NY: CLC Publications

Argument Asymmetry in German Cleft Sentences[*]

Swantje Tönnis, Lea M. Fricke, Alexander Schreiber

Georg-August-Universität Göttingen

Abstract. We present a corpus study on German *it*-clefts that tests whether subject clefts are more frequent than other clefts in German. This observation has been made for several other languages. However, we used a more complex method than earlier studies by not only providing the frequencies of (non-)subject clefts, but by additionally comparing those frequencies to the general frequency of (non-)subjects. Our results support the claim that subject clefts are more frequent in German. We argue that a cleft construction in its function to mark focus appears more often with subjects since there are additional options to mark focus on non-subjects. The importance of contrast, exhaustivity, and an existential presupposition as a motivation to use a cleft was also taken into account but did not turn out to be significant in our cleft sample. From these results, we conclude that subjecthood is the main factor that facilitates the use of a cleft, possibly as a result of the speaker's intention to give cues for prosodic prominence of an element.

Keywords: German it-clefts, prosodic prominence, focus, subject/non-subject asymmetry

1 Introduction

The aim of this paper is to contribute to a better understanding of the factors that facilitate the use of *it*-clefts in German. For this purpose, we conducted a corpus study in which we analyzed crucial properties of clefts and their contexts. In this paper, we mainly focus on one particular aspect, namely the grammatical role of the pivot. Depending on the grammatical role of the pivot in the relative clause, one can distinguish between subject clefts as in (1), object clefts as in (2), and PP clefts as in (3), among others. In this paper, we will consider two main groups: subject clefts and non-subject clefts, the latter containing either an object noun phrase or a prepositional phrase in the pivot.

(1) Es ist Peter, der Maria liebt.
It is Peter who$_{\text{NOM.SG}}$ Maria loves.
'*It is Peter who loves Maria.*'

[*] This research has been funded by the German Research Foundation (DFG) as part of the project "Exhaustivity in *it*-Clefts" in the priority program XPrag.de, which we thankfully acknowledge.

(2) Es ist Peter, den Maria liebt.
It is Peter who$_{\text{ACC.SG}}$ Maria loves.
'*It is Peter who Maria loves.*'

(3) Es ist Peter, zu dem Maria gegangen ist.
It is Peter, to who$_{\text{DAT.SG}}$ Maria gone is.
'*It is Peter who Maria went to.*'

It is a standard claim in the literature that subject clefts are more frequent than object clefts (see Carter-Thomas 2009, Reichle 2014 for French; Collins 1991, Roland et al. 2007 for English; Skopeteas and Fanselow 2010 comparing English, French, Georgian, and Hungarian). A more detailed discussion of this claim will be presented in Section 2. Our corpus study provides novel data on the distribution of subject and non-subject clefts in German, for which such studies do not yet exist. In contrast to earlier studies on other languages, we used a more fine-grained method in evaluating our data in order to account for the effect of the general distribution of subjects and non-subjects (see Section 3). In our discussion in Section 4, we address the question how clefting a subject/non-subject relates to prosodic prominence and whether there are other factors that motivate the use of a cleft construction. Section 5 presents some ideas about topic-comment clefts and Section 6 concludes.

2 Background

As mentioned, it is a standard claim that subject clefts are more frequent than object clefts. This observation is related to focus marking. Languages have different options to realize focus such as prosodic prominence, movement, morphology, or constructions like a cleft. However, not all of these options are equally available for all grammatical functions (see Lambrecht 2001 for French or Hartmann and Zimmermann 2007 for West Chadic Languages). In French, for example, prosodic prominence is a possible means to mark focus for objects, but not for subjects. Prosodic prominence is obligatorily realized at the right edge of the phonological phrase (see Féry 2001) and subjects cannot appear in that position. Objects, in contrast, do occur in this position and receive prosodic prominence. That does not mean that focus on an object cannot also be realized by a cleft construction. However, since there are other options for focus marking on objects as well, they are predicted to be clefted less often than subjects. A similar reasoning is formulated for other languages in Skopeteas and Fanselow (2010).

Also, Szendröi (1999, p. 553) proposes to analyze clefts in English as focus-driven movement. She argues that the subject is moved into an object position with the "dummy verb" *is* and the "dummy subject" *it*, where it receives the same default prominence as any other object (see also Reinhart 1995, p. 62). The default intonation of a focus-background cleft is exemplified in (4).

(4) Es ist PETER, der Maria liebt.
It is PETER who$_{\text{NOM.SG}}$ Maria loves.
'*It is PETER who loves Maria.*'

DeVeaugh-Geiss et al. (2015, p. 386) call clefts a structural device to mark focus unambiguously, similar to Hungarian pre-verbal focus (Szabolcsi 1981; É. Kiss 1998; Onea and Beaver 2009). However, they only claim this for focus-background clefts (see Section 4 and 5), which consist of a focused cleft pivot and a backgrounded existential presupposition cleft relative.

The aim of our study is to analyze the data with respect to the frequency of (non-)subject clefts and, thus, gain a deeper understanding of the function of a cleft sentence. More precisely, we discuss whether a cleft could serve as focus marking device.

3 Methods and Results

In our pilot corpus study, we tested whether subject clefts are more frequent than other clefts for written German, which to our knowledge has not been explicitly shown yet. However, we used a more complex method than earlier studies of other languages. We did not only provide the frequencies of subject and non-subject clefts, but also compared those frequencies to the general frequency of subjects and non-subjects. It is important to take this additional step since it could be possible that subjects are just clefted more often because they are generally more frequent.

We drew a sample of 300 random occurrences of clefts from the DeReKo corpus[1] of written German. We annotated[2] the grammatical function of the relative pronoun of each cleft relative. Moreover, we set up a comparison corpus of 200 non-clefted sentences from the same texts in which we found the clefts. Those sentences serve as a comparison in order to capture the frequency of certain grammatical categories in general. In Table 1, we present the absolute numbers n_{cleft} of (non-)subjects in the pivot of a cleft and the absolute numbers n_{comp} of (non-)subjects in non-clefted sentences. They seem to confirm that subject clefts are more frequent than non-subject clefts, also when compared to the frequency of subjects in the comparison corpus.

	n_{cleft}	n_{comp}
Subjects	249	191
Non-Subjects	51	274

Table 1. Absolute numbers n_{cleft} for cleft sample and n_{comp} for the comparison corpus.

It is not obvious how to interpret the observed frequencies in the cleft sample and in the comparison corpus. First of all, the general frequency of subjects in

[1] Das Deutsche Referenzkorpus DeReKo http://www.ids-mannheim.de/kl/projekte/korpora/, Institut für Deutsche Sprache, Mannheim

[2] The annotators were the three authors of this paper.

the comparison corpus plays a role (see (i) below), but it ignores the fact that only one cleft can be created from one CP at a time and various grammatical arguments are unevenly distributed in CPs. This way, we cannot account for the fact that in some clefts it is simply not possible to cleft a non-subject since it only contains a verb and a subject. Another approach (see (ii)) that does not suffer from this shortcoming is measuring the probability of each individual grammatical argument to become a cleft pivot. However, it rests on the idealized assumption that each CP is equally likely to become a cleft.

Each of these two approaches can be seen as a useful simplification because the aspects that they ignore are independent of each other. So it is safe to assume that subject clefts are more frequent if both approaches yield the same result. The two approaches are presented in the following:

(i) We determined the relative frequencies $\mathrm{f}_{\mathrm{comp}}$ of subjects and non-subjects in the comparison corpus by counting all of their occurrences in order to compare them to the observed relative frequencies $\mathrm{f}_{\mathrm{cleft}}$ in the cleft sample, ignoring their unequal distribution over the different types of sentences.

(ii) We calculated the probability to be clefted for each subject and non-subject in each sentence from the comparison corpus. Take an example sentence of the form S-O-V-PP-PP. The probability for the subject to be clefted in that sentence is 0.25. The probability that one of the other arguments is clefted is 0.75. Note that it is not possible to cleft the verb. Given a different sentence from the comparison corpus, say S-V, the probability for the subject to be clefted is 1. Then, we calculated the average $\mathrm{p}_{\mathrm{cleft}}$ of the probabilities of a subject/non-subject to be clefted over all sentences and compared them to the observed frequencies.

Both approaches yield that subject clefts occur significantly more often than non-subject clefts, even with respect to the general occurrence of subjects and non-subjects. For approach (i), we tested the relative frequencies $\mathrm{f}_{\mathrm{cleft}}$ of subjects and non-subjects from the cleft sample and the relative frequencies $\mathrm{f}_{\mathrm{comp}}$ from the comparison corpus for significant deviation using a χ^2-test. The frequencies are displayed in Table 2. The test shows that subject clefts are significantly more frequent in the cleft sample ($p<0.01$).

	$\mathrm{f}_{\mathrm{cleft}}$	$\mathrm{f}_{\mathrm{comp}}$
Subjects	0.83	0.41
Non-Subjects	0.17	0.59

Table 2. Results for approach (i).

For approach (ii), we tested $\mathrm{f}_{\mathrm{cleft}}$ and the average probabilities $\mathrm{p}_{\mathrm{cleft}}$ of subjects and non-subjects from the comparison corpus for significant deviation using a t-test. This test shows that subject clefts are significantly more frequent in the cleft sample than predicted by p_{cleft} ($p<0.01$). The frequencies and probabilities are displayed in Table 3.

	f_{cleft}	p_{cleft}
Subjects	0.83	0.51
Non-Subjects	0.17	0.49

Table 3. Results for approach (ii).

4 Discussion

The discussion will be concerned with focus-background clefts. Note that there are two types of *it*-clefts: focus-background clefts as in (5) and topic-comment clefts as in (6). We found 271 focus-background clefts and 28 topic-comment clefts, which shows that focus-background clefts are much more frequent. See Section 5 for some ideas about topic-comment clefts.

(5) John, Peter and Max were at the party. Somebody smoked and somebody danced. It was Max who smoked.

(6) John is a really good friend of mine. It was he who helped me when I was once very desperate.

For focus-background clefts, our observation can be explained in a similar way as mentioned above for French. In German, there are no prosodic constraints on the position of the element with the highest prominence. In spoken German, the context and intonation interact, as in (7). It is in general possible to mark any constituent in its base position by giving it prosodic prominence.

(7) Wer hat einen Apfel gegessen? – NINA hat einen Apfel gegessen.
Who has an apple eaten? – NINA has an apple eaten.
'*Who ate an apple? – NINA ate an apple.*'

For written German, however, the reader needs to infer the intonation from cues provided by the context or the sentence itself. If there are no contextual cues, the reader will assume the default intonation. In most cases, this leads to the highest prosodic prominence on the object as in (8). The object in (8) is the peripheral element within the prosodic domain, which Büring (2007) identifies as the focus position. Prepositional phrases, in contrast, do not always receive prosodic prominence as indicated in (9). However, they can be moved to that position by scrambling like in (10) (see den Besten 1983).[3]

[3] There is another option of moving an object or a prepositional phrase to the left periphery of the sentence as in (i).

(8) Laura hat ein BUCH gelesen.
Laura has a BOOK read.
'*Laura read a book.*'

(9) Laura hat in der Hängematte ein BUCH gelesen.
Laura has in the hammock a BOOK read.
'*Laura read a book in the hammock.*'

(10) Laura hat ein Buch in der HÄNGEmatte gelesen.
Laura has a book in the HAMMOCK read.
'*Laura read a book in the hammock.*'

Subject NPs, on the contrary, are neither prominent by default nor can they be scrambled or moved to a higher position (at least in most of the cases). Hence, there must be other cues for the reader to assume prosodic prominence on a subject. A cleft is such a cue. Again, clefting an object NP or a PP would also be a cue for marking the respective phrase with higher prosodic prominence. However, there are other options for object NPs and PPs that are not available for subject NPs. Hence, subject NPs are predicted to be more likely to be clefted than non-subjects.

There could be other factors such as agentivity and animacy – which often go along with subjecthood – that might account for the frequency of clefted subjects.[4] In this case, subjecthood might not be the crucial predictor for the higher frequency of subject clefts. When comparing the role of those categories in our cleft sample and in the comparison corpus, it did, however, turn out that they had no effect. If subjects were just clefted because they are agents or animate, we should find more agents or animate subjects in clefts than in the sentences from our comparison corpus. Table 4 displays the frequencies of (non-)animate subjects in clefts (f_{cleft}) and non-clefted sentences (f_{comp}). Table 5 displays the frequencies of (non-)agentive subjects in clefts (f_{cleft}) and non-clefted sentences (f_{comp}). We performed a χ^2-test for both categories, which showed no significant deviation at any level of significance (neither for animacy nor agentivity). Hence, subjecthood seems to be the crucial feature so far.

(i) In der HÄNGEmatte hat Laura ein Buch gelesen.
In the HAMMOCK has Laura a book read.
'*In the hammock, Laura read a book.*'

This kind of movement, however, is compatible with different accent patterns. At least, it is not obvious what the default intonation is. Example (i) could also receive the highest prominence on the object. Thus, this kind of movement is a less clear cue for predicting the highest prosodic prominence (see Frey 2004).

[4] Thanks to an anonymous reviewer for pointing this out.

	f_{cleft}	f_{comp}
subjects [+animate]	0.47	0.52
subjects [-animate]	0.53	0.48

Table 4. Animacy of subjects in the cleft sample and the comparison corpus.

	f_{cleft}	f_{comp}
subjects [+agent]	0.34	0.37
subjects [-agent]	0.67	0.63

Table 5. Agentivity of subjects in the cleft sample and the comparison corpus.

Hence, our results are in line with the proposal in DeVeaugh-Geiss et al. (2015), who claim that a cleft is a device to mark subject focus unambiguously which is not possible in the canonical word order in written German. Still, the question remains whether marking prosodic prominence and focus is the main motivation for a speaker to use a cleft sentence. Various other features of clefts have been mentioned in the literature, such as exhaustivity or an existential presupposition, which could motivate the use of a cleft. However, these features do not seem to be decisive for an appropriate use of a cleft. The status or even the existence of exhaustivity in clefts is hotly debated anyway. Horn (2014) cites the cleft in (11) from a poem by James Oppenheim in 1911, that does not support an exhaustivity inference of clefts.

(11) Yes, it is bread we fight for, but we fight for roses too!

For the categories exhaustivity and existential presupposition, the inter-annotator agreement was extremely low; i.e., these notions could not be made sufficiently operational for our corpus study. While we hope to remedy this in future research, we take this low agreement level as an preliminary indication that these categories might not be the main predictors for the usage of clefts.[5]

Another important aspect that supports our approach concerns the hypothesis about differences between spoken and written German. Based on native speaker intuition, we assume that clefts are much less frequently used in spoken German than in written German. We argue that an account just based on the existential presupposition and/or exhaustivity cannot explain those differences. There is no reason why the existential presupposition or the exhaustivity inference should behave differently in spoken as opposed to written German. Both features seem implausible to have an effect on the frequency of clefts in general. Our analysis of clefts as devices to shift prominence away from the default, in contrast, predicts there to be less clefts in spoken German. The reason is that

[5] Both categories are problematic for annotation purposes because annotators highly diverge in accommodation strategies, which affects the judgement for those categories in an unclear way.

there is always the option of marking any element in its base position just by means of intonation, as in (7). Hence, there is no need for a cleft construction in order to raise the prosodic prominence of a subject and disambiguate focus in spoken German.

Still, we do not challenge that clefts have an existential presupposition. Since we assume that the prosodic prominence corresponds to focus, an existential presupposition seems plausible to arise. Following Geurts and Van der Sandt (2004, p. 2), we take focus by itself to have an existential presupposition anyway. Thus, our analysis is not incompatible with clefts having precisely such a presupposition.

Another potential motivation for using a cleft is expressing contrast. For spoken language, Alter et al. (2001) and Sudhoff (2010), among others, state that contrast might be marked by a specific prosody different from focus. Since this prosody is not visible in written language, there must be other cues for contrast in written language. Destruel and Velleman (2014), among others, claimed that clefts fulfill this function.

In order to estimate the role of contrast for clefts, we annotated several other categories for our sample of cleft sentences (including the contexts) that could possibly relate to contrast. The notion of contrast itself is quite unclear in the literature. Repp (2010, p. 1335), for example, mentions the relevance of explicit alternatives for a contrastive focus as opposed to 'normal' focus. Therefore, we annotated the availability of explicit alternatives to the cleft pivot in the context. Furthermore, we checked whether there was a negation of the content expressed by the cleft in the context. In those cases, the cleft might be used as a correction which means that the cleft stands in contrast to the corrected part. However, those contrast related categories did not seem to play an important role for clefts in our sample.[6]

5 Topic-comment clefts

We distinguish between focus-background and topic-comment clefts. In contrast to focus-background clefts, it is commonly assumed that the pivot is not focused in topic-comment clefts since the cleft relative is the part that provides new information. This raises the question why a cleft, a structure that heightens the prominence of the pivot, is used at all. From our data, we can conclude that topic-comment clefts are, in fact, atypical clefts in that they occur rarely relative to the focus-background type. A reason to use them could lie in the circumstance that the discourse preceding the topic-comment cleft seems to feature the same topic over the course of several sentences. Topic-comment clefts might be used to re-heighten the prosodic prominence of the topic, possibly because the comment contains information of particular importance.

Furthermore, it is not surprising that topic-comment clefts are frequently subject clefts just like focus-background clefts, since topics in general are very

[6] The analysis of these categories in non-clefted sentences goes beyond the scope of this project.

likely to be subjects. Reinhart (1981, p. 62) states that there is a strong preference for the subject to be interpreted as the topic of a sentence, although it is just a tendency. Accordingly, whenever a cleft construction contains a topic in the pivot, it is very likely to be a subject.

Hence, our approach explains both types of clefts on a par – as devices to raise the prosodic prominence of the pivot. It is, however, more difficult to make predictions about differences between spoken and written German as far as topic-comment clefts are concerned. We have no data or a clear intuition about whether topic-comment clefts are used in spoken German at all. However, there is another option to re-heighten the prosodic prominence that is frequently used in spoken language, namely the left dislocation as in (12) (see Reinhart 1981, p. 63), but this strategy seems to be less frequent in written language.

(12) Felix, it's been ages since I've seen him.

Reinhart (1981, p. 63) argues that this movement marks the moved element as the topic in most cases. It is not yet clear to us how this interacts with the acceptability and the frequency of topic-comment clefts in spoken German. If left dislocation was a good alternative to using a topic-comment cleft in spoken German, our approach would predict there to be less topic-comment clefts in spoken German. If, on the contrary, clefting is just a device of re-heightening the prominence of a topic and left-dislocation is not an option, we predict topic-comment clefts to appear in spoken language as often as in written language. Different from focus, this re-heightening cannot be achieved by intonation. It is not clear whether there is a specific intonation that is used for or perceived as re-heightening.

6 Conclusion

From our data set, we can so far only conclude that subjecthood is the main factor determining the use of clefts, possibly due to the wish of the speaker to give cues for the prosodic prominence of the argument functioning as subject, which is not marked for prominence by default. Accordingly, clefts can be assumed to constitute a strategy to disambiguate focus marking in written German by moving the pivot into a position with default high prominence. This analysis predicts subject clefts to be more frequent than non-subject clefts, since German has other ways of making an object or a prepositional phrase prominent, e.g., default intonation and movement.

Still, more detailed research is needed as far as the contrast categories are concerned. It is not trivial how to define contrast related categories for the purpose of annotation since the notion of contrast is debated anyway. Also, one should look into the differences between spoken and written language and address the question why clefts are used more frequently in written German and provide empirical studies on the frequency of *it*-clefts in spoken German. If clefts do give cues about intonation, the analysis of spoken data could be very fruitful.

Bibliography

Alter, K., Mleinek, I., Rohe, T., Steube, A., and Umbach, C. (2001). Kontrastprosodie in Sprachproduktion und -perzeption. *Linguistische Arbeitsberichte*, 77:59–79.

Büring, D. (2007). Semantics, intonation, and information structure. In Ramchand, G. and Reiss, C., editors, *The Oxford Handbook of Linguistic Interfaces*, pages 445–474. Oxford University Press, Oxford.

Carter-Thomas, S. (2009). The french c'est-cleft: Function and frequency. In Banks, D., editor, *La linguistique systémique fonctionnelle et la langue française*, pages 127–157. L'Harmattan, Paris.

Collins, P. C. (1991). Pseudocleft and cleft constructions: A thematic and informational interpretation. *Linguistics*, 29(3):481–520.

den Besten, H. (1983). On the interaction of root transformations and lexical deletive rules. In Abraham, W., editor, *On the formal syntax of the Westgermania*, pages 47–131. John Benjamins Publishing Company, Amsterdam, Philadelphia.

Destruel, E. and Velleman, L. (2014). Refining contrast: Empirical evidence from the english it-cleft. *Empirical Issues in Syntax and Semantics*, 10:197–214.

DeVeaugh-Geiss, J., Zimmermann, M., Onea, E., and Boell, A.-C. (2015). Contradicting (not-)at-issueness in exclusives and clefts: An empirical study. *Semantics and Linguistic Theory*, 25:373–393.

É. Kiss, K. (1998). Identificational focus versus information focus. *Language*, 74(2):245–273.

Féry, C. (2001). Focus and phrasing in french. In Fery, C. and Sternefeld, W., editors, *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow*, pages 153–181. Akademie Verlag, Berlin.

Frey, W. (2004). The grammar-pragmatics interface and the German prefield. *Sprache un Pragmatik*, 52:1–39.

Geurts, B. and Van der Sandt, R. (2004). Interpreting focus. *Theoretical Linguistics*, 30:1–44.

Hartmann, K. and Zimmermann, M. (2007). In place-out of place? Focus in Hausa. In Schwabe, K. and Winkler, S., editors, *On Information Structure, Meaning and Form: Generalizations across languages*, pages 365–403. John Benjamins Publishing Company, Amsterdam, Philadelphia.

Horn, L. (2014). Information structure and the landscape of (non-)at-issue meaning. In Féry, C. and Ishihara, S., editors, *The Oxford Handbook of Information Structure*. Oxford University Press.

Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics*, 39(3):463–516.

Onea, E. and Beaver, D. (2009). Hungarian focus is not exhausted. In *Proceedings of Semantics and Linguistic Theory (SALT) 19*, pages 342–359.

Reichle, R. V. (2014). Cleft type and focus structure processing in French. *Language, Cognition and Neuroscience*, 29(1):107–124.

Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27(1):53–94.

Reinhart, T. (1995). Interface strategies. *OTS working papers in linguistics.*

Repp, S. (2010). Defining 'contrast' as an information-structural notion in grammar. *Lingua*, 120(6):1333–1345.

Roland, D., Dick, F., and Elman, J. L. (2007). Frequency of basic english grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3):348–379.

Skopeteas, S. and Fanselow, G. (2010). Focus types and argument asymmetries: a cross-linguistic study in language production. In Breul, C. and Göbbel, E., editors, *Contrastive information structure*, pages 169–197. Benjamins, Amsterdam, Philadelphia.

Sudhoff, S. (2010). Focus particles and contrast in German. *Lingua*, 120(6):1458–1475.

Szabolcsi, A. (1981). The semantics of topic-focus articulation. In Groenendijk, J. A. G., Janssen, T. M. V., and Stokhof, M. B. J., editors, *Formal methods in the study of language*, pages 513–540. Mathematisch Centrum, Amsterdam.

Szendröi, K. (1999). A stress-driven approach to the syntax of focus. *UCL Working Papers in Linguistics*, 11:545–573.

Revisiting the secrets of BEFORE: lessons from Modern Greek[⋆]

Orest Xherija

Department of Linguistics, The University of Chicago, Chicago IL 60637, U.S.A.
orest.xherija@uchicago.edu

Abstract. I consider two analyses of the semantics of BEFORE-clauses (**BC**s) in light of two phenomena in Modern Greek (**MG**): licensing of strong Negative Polarity Items (NPIs) and an anti-PAST restriction on the verb in the **BC**. I show that [2] and [11] cannot be extended to **MG** (at least without significant modifications) and that a new approach is necessary. This paper proposes a disjunctive semantics for BEFORE that makes **BC**s non-committal by default (that is, there is no commitment about the instantiation of the event described by the **BC**) and makes the factual and non-factual inferences contextual entailments. The disjunctive semantics makes BEFORE a NONVERIDICAL environment which explains the licensing of weak NPIs in **BC**s and the emergence of the PERFECTIVE NON-PAST (PNP) as the tense-aspect combination of the verb of **BC**s. The licensing of strong NPIs is achieved through a rescuing mechanism similar to that of [6].

1 Introduction

It is a well-attested fact of English that BEFORE-clauses (**BC**s) can yield a factual [1], a non-factual [2] and a non-committal [3] inference about the instantiation of the eventuality they describe.

1. Dreyfus ate the salad BEFORE he had dessert.
   $\Longrightarrow$ Dreyfus had desert. (**factual**)
2. The MI6 defused the bomb BEFORE it exploded.
   $\Longrightarrow$ The bomb did **not** explode. (**non-factual**)
3. Dreyfus left the country BEFORE anything **ever** happened.
   $\not\Longrightarrow$ Something did (not) happen. (**non-committal**)

A natural question is whether, crosslinguistically, words whose meaning is akin to that of English BEFORE, namely words which (at least in an intuitive sense) are used to talk about temporal precedence, exhibit similar semantic behavior. It turns out that these patterns are crosslinguistically robust and can

[⋆] I am grateful to Anastasia Giannakidou, Itamar Francez, Alda Mari, Chris Kennedy, Cleo Condoravdi, Yael Sharvit and Henriëtte de Swart for valuable comments and suggestions. Naturally, the aforementioned individuals do not necessarily agree with the views of this paper and the responsibility for all errors remains solely with me.

be observed in a number of languages, including Italian [3], German [12,15], Catalan [13], Russian [15] and Japanese [10,9, *inter alia*]. The following examples, which are direct translations of [1] - [3] in Modern Greek (**MG**), show that the English inference pattern is observed in this language, too:

4. O Dreyfus éfaɣe ti saláta PRIN fái to ɣlikó.
$\Longrightarrow$ O Dreyfus éfaɣe to ɣlikó. (**factual**)
5. I MI6 apenerɣopíise ti vómva PRIN ekraɣí.
$\Longrightarrow$ I vómva **ðen** ekseráɣi. (**non-factual**)
6. O Dreyfus éfiɣe apó ti χóra PRIN simví **poté** típota.
$\not\Longrightarrow$ Káti (ðen) sinévi. (**non-committal**)

A second robust crosslinguistic fact is that BEFORE licenses weak Negative Polarity Items (NPIs) in the **BC**, as the presence of *ever* in the **BC** of [3] and of *poté* 'ever' in the **BC** of [6] exemplify. In this paper, I want to consider two phenomena from **MG BC**s that, to the best of my knowledge, have not been addressed in the literature and their study might shed light to some intricacies in the meaning of **BC**s.

- **MG BC**s sporadically allow strong NPIs à la [18], that is NPIs that need to be in the scope of an at least ANTIADDITIVE operator, as exemplified by the presence of focused *kanéna* in the **BC** of [7]; and
- they forbid PAST tense marking on their verb and only allow it to surface in the PERFECTIVE NON-PAST (PNP) form [8], a tense-aspect combination that is only sanctioned in NONVERIDICAL contexts[1] in **MG**, as argued in [7]. This does not hold true for other **MG** temporal connectives as can be seen in[2] [8], where AFTER- and WHEN-clauses do not forbid PAST tense marking on the verb.

7. O Iorðánis péθane PRIN ði / *íðe **kanéna**$_{\text{F}}$ egóni tu.
the Jordan died BEFORE see.PNP / saw **nobody** grandchild his
'Jordan died before seeing **any at all** of his grandchildren.'
8. I Féðra éfiɣe ÓTAN/AFÚ *ftási / éftase i Natasa.
the Phaedra left WHEN/AFTER arrive.PNP / arrived the Natasha
'Phaedra left when/after Natasha arrived.'

This paper aims to address three questions: (a) How do the inferences in [1] – [3] arise and what is their truth-conditional status? (b) How is the PNP verbal form in **MG** related to the potential (non)veridicality of BEFORE? and (c) How does the licensing of (strong) NPIs take place in **MG BC**s?

[1] An operator $\mathscr{F}$ is NONVERIDICAL if for all propositions p, $\mathscr{F}(p) \not\Longrightarrow p$.

[2] Some **MG** temporal connectives are followed by certain particles that impose their own selectional restrictions on the verb. I do not address this class of temporal connectives in this paper.

2 Previous work

The most successful analyses of the meaning of **BC**s are by [2], who develops an intensional account for temporal clauses (and **BC**s in particular), and [11], who provides a Gricean account of the relevant phenomena. It can be shown that these accounts cannot be extended to model the **MG** data we presented. The intentional account of [2] relies on STRAWSON DOWNWARD ENTAILMENT (SDE) to account for NPI-licensing in **BC**s, which has been shown to face challenges with **MG** NPIs [6] across the board, not only in temporal clauses. To illustrate one shortcoming of the SDE approach to NPI-licensing in **MG** **BC**s, consider the following sentences:

9. Páre **kanéna** milo.
   take.IMP.2SG any.NPI apple.
   'Take some apple or other.' (non-SDE; NPI licensed)

We note that SDE is not even a necessary condition for NPI-licensing, since imperatives are not SDE environments but still license NPIs in **MG** as illustrated in [9]. Imperatives are not the only non-SDE environments that license NPIs in **MG**. Future tense, modals and exclusive disjunction are some other non-SDE operators that license NPIs (see [5] for a thorough distribution of **MG** NPIs). [4]'s SDE approach is not able to handle the distribution of **MG** NPIs. In [2]'s approach, **BC**s support strengthening inferences in terms of STRAWSON ENTAILMENT and therefore create SDE contexts but this approach will not do the trick for the NPIs we consider.

Turning to the Gricean account, one observes that it employs a denotation of BEFORE that renders it ANTIADDITIVE[3], and according to [17] predicts licensing of strong NPIs in all **BC**s, a prediction that does not hold for many languages as exemplified by [10] and [11].

10. *I Avɣeriní éfaɣe mesimerjanó PRIN meletísi **kanéna**$_{\mathrm{F}}$ máθima.
    the Avgerini ate lunch BEFORE study.PNP none lesson
    (MG)
11. *Lira iku PARA se të shikonte **asnjërin**.
    Lira left BEFORE than SUBJ see.SUBJ.3SG nobody
    (Albanian)

A more detailed discussion of these (and other) shortcomings of these accounts has to be postponed due to space limitations but I hope to have convinced the reader that an immediate application (that is, without significant modifications) of either of these two accounts to **MG** would be unsuccessful.

[3] A proof of the ANTIADDITIVITY of [11]'s before is in Appendix [A].

3 Proposal

I restrict my attention to BEFORE when it conjoins two untensed clauses; I ignore BEFORE with a nominal complement. I take it, following [14, among others], that verbs require a time-interval argument of the form $[a,b]$, $a \prec b$. The type of time intervals will be i and therefore the type of temporal properties will be $\langle i,t\rangle$. I do not take any position regarding the properties of the denotation of verbs depending on their *Aktionsart* class, but the reader can consult [2] for a possible set of assumptions. We assume, with [2], that the untensed clause $[\mathscr{A}$ BEFORE $\mathscr{B}]$ composes intersectively, i.e. $\llbracket \mathscr{A}\ \text{BEFORE}\ \mathscr{B} \rrbracket = \llbracket \mathscr{A} \rrbracket \wedge \llbracket \text{BEFORE}\ \mathscr{B} \rrbracket$. Finally, we denote by "$\prec$" the relation of temporal precedence and by "inf" the greatest lower bound of a non-empty set of $\mathbb{R}$, with the additional premise that there exists an isomorphism between $\mathbb{R}$ and the set of moments of time $\mathscr{T}$. With this background, we propose the denotation for BEFORE in [B0], where $\veebar$ is exclusive disjunction:

$$\llbracket \text{BEFORE} \rrbracket = \lambda \mathscr{X}_{\langle i,t\rangle} \lambda t_i \Big[\Big((\exists t'' \neq \emptyset) \big[(\inf(t) \prec \inf(t'')) \wedge \mathscr{X}(t'') \big] \Big) \veebar \Big(\forall t' \big[\neg \mathscr{X}(t') \big] \Big) \Big] \tag{B0}$$

$$\llbracket \text{BEFORE}\ \mathscr{B} \rrbracket = \lambda t_i \Big[\Big((\exists t'' \neq \emptyset) \big[(\inf(t) \prec \inf(t'')) \wedge \mathscr{B}(t'') \big] \Big) \veebar \Big(\forall t' \big[\neg \mathscr{B}(t') \big] \Big) \Big] \tag{B1}$$

As a temporal property, [B1] can intersectively combine with $\mathscr{A}$ to yield the truth conditions in [B2]:

$$\llbracket \mathscr{A}\ \text{BEFORE}\ \mathscr{B} \rrbracket = \lambda t_i \Big[\underline{\mathscr{A}(t) \wedge \Big(\Big((\exists t'' \neq \emptyset) \big[(\inf(t) \prec \inf(t'')) \wedge \mathscr{B}(t'') \big] \Big) \veebar \Big(\forall t' \big[\neg \mathscr{B}(t') \big] \Big) \Big)} \Big] \tag{B2}$$

Under the simplifying assumption that there is one PAST tense operator scoping above both clauses and denoting the underlined portion of [B2] by $\mathscr{E}$, the utterance time by t_{UT}, the contextually restricted relevant time interval by $\mathscr{T}_c$ and the least upper bound of a set of $\mathbb{R}$ by "sup" we obtain the truth conditions in [B3]:

$$\llbracket \text{PAST} \rrbracket \Big(\llbracket \mathscr{A}\ \text{BEFORE}\ \mathscr{B} \rrbracket \Big) = \exists t \subset \mathscr{T}_c \Big(\big(t \neq \emptyset \wedge \sup(t) \preceq t_{\text{UT}}\big) \wedge \mathscr{E} \Big) \tag{B3}$$

Informally, this approach, similar in spirit to [11], claims that a sentence $[\mathscr{A}$ BEFORE $\mathscr{B}]$ is true either if event $\mathscr{B}$ occurs at a time after $\mathscr{A}$ or if it is not instantiated at all in the contextually relevant interval.

4 The nature of the inferences

The default inference is the non-committal. More specifically, in out-of-the-blue contexts, i.e. in situations in which there is no discourse-specific information added to the CONTEXT, the exclusive disjunction does not allow resolution in favour of any of the two disjuncts. The factual and non-factual inferences arise as contextual entailments from the disjunction elimination rule [DE] below:

$$\frac{\mathscr{X} \veebar \mathscr{Y} \quad \neg\mathscr{X}}{\mathscr{Y}} \vee E \tag{DE}$$

The motivation for this is apparent. **BC**s are disjunctive propositions, so if the CONTEXT contains the negation to one of the disjuncts of a **BC**, the remaining disjunct will be the contextually entailed one. In particular, if the meaning of the **BC** is $\mathscr{A} \veebar \mathscr{B}$ and we can deduce $\neg\mathscr{B}$ (respectively $\neg\mathscr{A}$) from the set of premises containing the common ground and the main clause with its presuppositions and entailments, then by [DE], $\mathscr{A}$ (respectively $\mathscr{B}$) can be concluded. In [7], Jordan dying has an entailment that he cannot be the agent of any action occuring after the time of death. This entailment together with disjunction elimination contextually entails the negative disjunct in the denotation of [$\mathscr{A}$ BEFORE $\mathscr{B}$], namely that Jordan did not see his grandchildren.

In an analogous fashion, one derives the positive disjunct from contexts that favour it. Consider [12] below:

12. **Q:** When did John wash his car?
    A: BEFORE he mowed his lawn.

If we assume that *wh*-adjunct questions carry an existential presupposition (following work such as [8] and [1] *inter multa alia*), then the expected answer to the question will be a time specification for the car-washing event. This presupposition of existence is the negation of the disjunct stating that "$\forall t' \left[\neg[\![\mathscr{B}(t')]\!]\right]$". Consequently, using ∨E we can conclude that the other disjunct is true.

There is one additional, typological observation that seems to favor an account in which BEFORE is by default non-committal. In **MG**, the verb of the **BC** is in a dependent form, as mentioned in the introduction. More precisely, it is in PERFECTIVE NON-PAST, a form that as [7] argues, "contains a dependent time variable, i.e. a referentially deficient variable that cannot be identified with the utterance time of the context". This restriction is only present for **BC**s, and does not surface with other temporal connectives. This referential deficiency of the PNP might serve as additional evidence for an ignorance-based account, such as the one I am advocating here.

5 The PNP verbal form

The PNP form of the verb is a weak NPI, per [5], as its presence is parasitic to that of a NONVERIDICAL environment. In particular, it is dependent on the presence of a subclass of NONVERIDICAL environments: the future, the subjunctive, the conditional and the optative. NONVERIDICALITY, however, is merely a necessary condition for the licensing of the PNP. For example, NEGATION, a prototypical NONVERDICAL operator does not license the PNP. This is because of selectional restrictions and additional semantic requirements of the PNP, thoroughly discussed in [7].

6 NPI-licensing

The denotation of BEFORE contains (exclusive) DISJUNCTION, a NONVERIDICAL operator, so adopting the theory of NPI-licensing of [5], which states that weak NPIs need to appear in NONVERIDICAL environments, we can see how examples like [3] are accounted for. Interestingly, exclusive disjunction does sanction weak NPIs in **MG** [13]:

13. I bíke **kanénas** sto spíti i afísame ta fota aniχtá.
or entered.3SG **anyone** at.the house or left.1PL the lights switched-on.PL
'Either **someone or other** entered the house or we left the lights on.'

For the licensing of the strong NPI in [7], we posit a rescuing mechanism in the spirit of [6]'s rescuing mechanism for explaining the occurrence of *any* under ONLY. We posit that strong NPIs are sanctioned in the presence of strictly nonveridical operators (that is, nonveridical but not antiveridical) if a negative inference is contextually entailed.

I want to conclude the discussion about strong NPI-licensing in **BC**s by briefly mentioning the results of [16]. [16] investigate the time course of processing negation by studying how the NPI *ever* is processed in different types of negative environments. Their results show that negative information from both asserted and non-asserted content, i.e. explicit and implicit negation, is accessed equally rapidly in online processing. However, they find that explicit negation, namely negation that is present in the syntactic-semantic representation is applied immediately to license NPIs while implicit or pragmatically inferred negation is adopted at a later processing stage as a last-resort NPI-licensing mechanism, leading to additional pragmatic processing cost. This is a potentially welcome result for the STRONG RESCUING hypothesis as it might be the case that an analogous mechanism is at play for the licensing of strong NPIs in **BC**s. Further experimental work is necessary to validate this hypothesis and will be the focus of future work.

7 Conclusion

This paper has reconsidered two analyses of the semantics of **BC**s in light of two phenomena in **MG BC**s : licensing of strong NPIs and the anti-PAST restriction on the verb. I showed that [2] and [11] cannot be extended to **MG** (at least without modifications) and that a new approach is necessary. The proposal in this paper proposes a disjunctive semantics for BEFORE that makes **BC**s non-committal by default and renders the factual and non-factual inferences contextual entailments The disjunctive semantics makes BEFORE a NONVERIDICAL environment and explains the licensing of weak NPIs in **BC**s and the emergence of the PNP as the tense-aspect combination of the verb of **BC**s. The licensing of strong NPIs is achieved through a rescuing mechanism similar to that of [6].

This paper is a small addition to the important literature about temporal clauses in particular, and adjunct clauses more generally. It enriches the verbal typology as far as verbal forms appearing in adjunct clauses are concerned and it adds to the long-standing problem of the semantics of BEFORE by taking crosslinguistic perspective. Finally, it adds to the vast literature on NPI-licensing byby calling attention to another potential mode of NPI-licensing, a licensing of last resort similar to that introduced in [6]

References

1. Comorovski, I.: Interrogative Phrases and the Syntax-Semantics Interface, Studies in Linguistics and Philosophy, vol. 59. Springer Netherlands (February 1996)
2. Condoravdi, C.: NPI licensing in temporal clauses. Natural Language & Linguistic Theory 28(4), 877–910 (November 2010)
3. Del Prete, F.: A non-uniform semantic analysis of the Italian temporal connectives *prima* and *dopo*. Natural Language Semantics 16(2), 157–203 (June 2008)
4. von Fintel, K.: NPI Licensing, Strawson Entailment, and Context Dependency. Journal of Semantics 16(2), 97–148 (1999)
5. Giannakidou, A.: Polarity Sensitivity as (Non)veridical Dependency, Linguistik Aktuell, vol. 23. John Benjamins Publishing Company (1998)
6. Giannakidou, A.: Only, emotive factive verbs, and the dual nature of polarity dependency. Language 82(3), 575–603 (September 2006)
7. Giannakidou, A.: The dependency of the subjunctive revisited: temporal semantics and polarity. Lingua 119(12), 1883–1908 (2009)
8. Karttunen, L., Peters, S.: What indirect questions conventionally implicate. In: Mufwene, S.S., Walker, C.A., Steever, S.B. (eds.) Papers from the Regional Meeting of the Chicago Linguistic Society. vol. 12, pp. 351–368. Chicago Linguistic Society (August 1976)
9. Kaufmann, S., Miyachi, M.: On the temporal interpretation of japanese temporal clause. Journal of East Asian Linguistics 20(1), 33–76 (2011)
10. Kaufmann, S., Takubo, Y.: Non-veridical uses of japanese expressions of temporal precedence. In: McGloin, N.H., Mori, J. (eds.) Japanese/Korean Linguistics. vol. 15. CSLI Publications (September 2007)
11. Krifka, M.: *Before* and *After* without coercion: comment on the paper by Cleo Condoravdi. Natural Language & Linguistic Theory 28(4), 911–929 (November 2010)
12. Manfred Krifka: How to interpret “expletive" negation under *bevor* in German. In: Hanneforth, T., Fanselow, G. (eds.) Language and Logos, studia grammatica, vol. 72, pp. 214–236. Akademie Verlag (2010)
13. Maria Teresa Espinal: Expletive Negation, Negative Concord and Feature Checking. In: Catalan Working Papers in Linguistics, vol. 8, pp. 47–69. Universitat Autònoma de Barcelona (2000)

14. Sharvit, Y.: On the universal principles of tense embedding: the lesson from *before*. Journal of Semantics 31(2), 263 – 313 (April 2014)
15. von Stechow, A., Grønn, A.: Tense in adjuncts part 2: Temporal adverbial clauses. Language and Linguistics Compass 7(5), 311–327 (May 2013), http://dx.doi.org/10.1111/lnc3.12019
16. Xiang, M., Grove, J., Giannakidou, A.: Semantic and pragmatic processes in the comprehension of negation. Journal of Neurolinguistics 38, 71–88 (2015)
17. Zwarts, F.: Nonveridical Contexts. Linguistic Analysis 25(3–4), 286–312 (1995)
18. Zwarts, F.: Three Types of Polarity. In: Hamm, F., Hinrichs, E. (eds.) Plurality and Quantification, Studies in Linguistics and Philosophy, vol. 69, pp. 177–238. Kluwer Academic Publishers (1998)

A [11]'s BEFORE is ANTIADDITIVE

Theorem 1. *Let* BEFORE *be defined as in [11]. Then* BEFORE *is* ANTIADDITIVE.

Proof. Let $\mathscr{B}, \mathscr{C}$ be arguments of BEFORE and denote by $\llbracket \mathscr{X}(t') \rrbracket^{t' \leq t}$ the expression $\left[t' \leq t \wedge \llbracket \mathscr{X}(t') \rrbracket\right]$. Recall, also, the following statements from propositional logic and set theory, where α denotes an arbitrary type:

$$
\begin{aligned}
&1.\ \neg(\exists x)[\mathscr{P}(x)] \equiv (\forall x)\neg[\mathscr{P}(x)] && \text{(NE)}\\
&2.\ \lambda x_\alpha.\big(\mathscr{X} \vee \mathscr{Y}\big) \equiv \lambda x_\alpha.\mathscr{X} \vee \lambda x_\alpha.\mathscr{Y} && \text{(PD)}\\
&3.\ \lambda x_\alpha.\big(\mathscr{X} \wedge \mathscr{Y}\big) \equiv \lambda x_\alpha.\mathscr{X} \wedge \lambda x_\alpha.\mathscr{Y} && \text{(PC)}\\
&4.\ \mathscr{A} \wedge (\mathscr{B} \vee \mathscr{C}) \equiv (\mathscr{A} \wedge \mathscr{B}) \vee (\mathscr{A} \wedge \mathscr{C}) && \text{(STA)}\\
&5.\ (\forall x)[\mathscr{P}(x) \wedge \mathscr{R}(x)] \equiv (\forall x)[\mathscr{P}(x)] \wedge (\forall x)[\mathscr{R}(x)] && \text{(QD)}
\end{aligned}
$$

Then:

$$
\begin{aligned}
&\llbracket \textsc{before}(\mathscr{B} \vee \mathscr{C}) \rrbracket\\
&\equiv \lambda t.\Big(\neg(\exists t')\llbracket(\mathscr{B} \vee \mathscr{C})(t')\rrbracket^{t' \leq t}\Big) && (\textbf{??})\\
&\equiv \lambda t.\Big((\forall t')\neg\llbracket(\mathscr{B} \vee \mathscr{C})(t')\rrbracket^{t' \leq t}\Big) && \text{(NE)}\\
&\equiv \lambda t.\Big((\forall t')\neg\Big[(t' \leq t) \wedge \big(\llbracket\mathscr{B}(t')\rrbracket \vee \llbracket\mathscr{C}(t')\rrbracket\big)\Big]\Big) && \text{(PD)}\\
&\equiv \lambda t.\Big((\forall t')\neg\Big[\llbracket\mathscr{B}(t')\rrbracket^{t' \leq t} \vee \llbracket\mathscr{C}(t')\rrbracket^{t' \leq t}\Big]\Big) && \text{(STA)}\\
&\equiv \lambda t.\Big((\forall t')\Big[\big(\neg\llbracket\mathscr{B}(t')\rrbracket^{t' \leq t}\big) \wedge \big(\neg\llbracket\mathscr{C}(t')\rrbracket^{t' \leq t}\big)\Big]\Big) && \text{(de Morgan)}\\
&\equiv \lambda t.\Big[\Big((\forall t')\neg\llbracket\mathscr{B}(t')\rrbracket^{t' \leq t}\Big) \wedge \Big((\forall t')\neg(\llbracket\mathscr{C}(t')\rrbracket^{t' \leq t}\Big)\Big] && \text{(QD)}\\
&\equiv \lambda t.\Big[(\forall t')\neg\Big(\llbracket\mathscr{B}(t')\rrbracket^{t' \leq t}\Big)\Big] \wedge \lambda t.\Big[(\forall t')\neg\Big(\llbracket\mathscr{C}(t')\rrbracket^{t' \leq t}\Big)\Big] && \text{(PC)}\\
&\equiv \lambda t.\Big[\neg(\exists t')\Big(\llbracket\mathscr{B}(t')\rrbracket^{t' \leq t}\Big)\Big] \wedge \lambda t.\Big[\neg(\exists t')\Big(\llbracket\mathscr{C}(t')\rrbracket^{t' \leq t}\Big)\Big] && \text{(NE)}\\
&\equiv \llbracket \textsc{before}(\mathscr{B}) \wedge \textsc{before}(\mathscr{C}) \rrbracket && (\textbf{??})
\end{aligned}
$$

$\therefore$ BEFORE is anti-additive.