SUPPLEMENTARY MATERIAL

1 FROM EVENT REPRESENTATIONS TO SBML

Existing NLP systems often use an event representation format comprised of a set of annotation rules and file formats to represent pathway events and entities (Kim *et al.*, 2011). For the purpose of this paper we base ourselves on the standoff representation (ST) proposed for the BioNLP Shared Task 2011, 2013 (Nédellec *et al.*, 2013).

Annotations in ST link spans of texts through character offsets to `entities` (e.g. Proteins, Genes etc.) and `events` (Positive Regulation etc.). Events and entities are represented line by line with links between them.

The following is an example sentence and a possible event representations (see also Figures 1 and 2).

(1) YAP modulates the phosphorylation of Akt1.

```
T1 Protein 0 3 YAP
T2 Protein 37 41 Akt1
T3 Regulation 4 13 modulates
T4 Phosphorylation 17 32 phosphorylation
E1 Phosphorylation:T4 Theme:T2
E2 Regulation:T3 Theme:E1 Cause:T1
```

Each annotation starts with a unique annotation-ID. The annotations-IDs encodes the annotation type in the first letter (T - text bound annotation, E - event annotation). This is followed by the annotation-type. For instance, the text bound annotation T1 is of type protein, whereas T3 is of type Regulation. Text bound annotations also encode the start and end position as well as the text they annotate. Text bound annotation T1 for instance ranges from character 0 to character 3 of the annotated text and the actual text is "YAP".

Event annotations build on top of text bound annotation. The annotations-ID for an event is followed by an event-type and the reference to the text bound annotation. For instance, E1 is a Phosphorylation event and the corresponding text is T4 "phosphorylation". Additionally, event annotations encode roles. T2 is the theme of E1, which in this case means that "Akt1" is undergoing a phosphorylation. Events can also be used as theme. For example the theme of E2 is E1, which means that the phosphorylation is regulated by "YAP". Different roles are possible depending on the type of the event.

Systems Biology Markup Language, or short SBML (Hucka *et al.*, 2003), is a XML-based markup language to describe, store and communicate biological models. It is among the most widely used formats with numerous software support. SBML essentially encodes models using biological players called `sbml:species`[1]. `sbml:species` participate in `sbml:reaction` interactions as `sbml:reactant`, `sbml:product` and `sbml:modifier`. The basic idea being that some quantity of reactant is consumed to produce a product. Reactions are influenced by modifiers.

SBML supports mathematical representations of the underlying dynamics of the reactions and is essentially used to simulate models. Due to this, there is no SBML vocabulary to specify different types of reactions (such as transcription, phosphorylation etc.) or species (such as protein, DNA etc.). Alternatively, species and reactions can be annotated and uniquely specified using MIRIAM resources and annotations (Novere *et al.*, 2005). We use controlled vocabulary from the Systems Biology Ontology (SBO) and the Gene Ontology (GO). This information is also useful to convert SBML files to other formats such as SBGN (Le Novere *et al.*, 2009) using tools such as VANTED (Junker *et al.*, 2006).

Figure 2 shows Example 1 converted into an SBML model using the mapping algorithm described in the following paragraphs.

[1] We will refer to SBML vocabulary using the prefix "sbml".

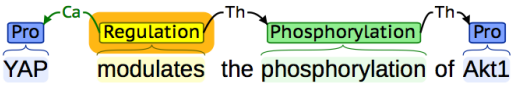

Fig. 1. Graphical representation of the event representation of Example 1

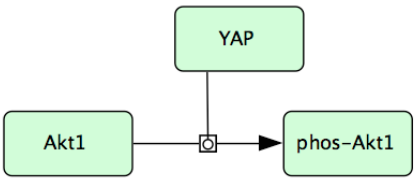

Fig. 2. Example 1 converted into SBML (viewed with CellDesigner)

Standoff Entity	SBO term	SBO name
Complex	SBO:0000253	non-covalent complex
Gene_or_gene_product	SBO:0000245	macromolecule
Dna	SBO:0000251	deoxyribonucleic acid
DnaRegion	SBO:0000251	deoxyribonucleic acid
Drug	SBO:0000247	simple chemical
Ion	SBO:0000327	non-macromolecular ion
Protein	SBO:0000252	polypeptide chain
Rna	SBO:0000250	ribonucleic acid
RnaRegion	SBO:0000250	ribonucleic acid
Gene	SBO:0000354	informational molecule segment
Small Molecule	SBO:0000247	simple chemical
Simple_molecule	SBO:0000247	simple chemical

Table 1. Mapping of entity trigger type to SBO term.

1.1 Mapping Algorithm

The conversion of standoff formatted information to an SBML model consists of five steps.

Step 1: Initialize the Model Firstly, read the event annotation files and create a memory internal representation of triggers and events. We initialize an empty SBML model with a single `sbml:compartment` named "default".

Step 2: Create **sbml:species** For each entity in the standoff format, a `sbml:species` is added to the SBML model. This only applies to standoff entities that can be mapped to an SBO term. Then the following is done 1) map the annotation-ID of

the trigger to the id in the `sbml:species`, 2) create a meta id by appending "metaid_0000" and annotation-ID; meta id facilitates that annotations to this species can uniquely refer to it 3) add the annotation-text as the name of the `sbml:species`, 4) map the annotation-type to an SBO term and add to the `sbml:species` (see Table 1)

For instance, the standoff line

```
T2 Protein 37 41 Akt1
```

will be mapped to

```
<species sboTerm="SBO:0000252"
id="T2" name="Akt1"
metaid="metaid_0000T2"
compartment="default"/>
```

On the other hand, a line such as

```
T39  Entity 641 648 nucleus
```

will not be used to create a species in the SBML model, because "Entity" cannot be mapped to an SBO term. Here, "nucleus" actually refers to a compartment which is not directly deducible from the entity definition in the standoff format. To deal with such cases, we need to take into account their role in Events something that is described in the next few paragraphs.

Step 3: Create **sbml:reaction** Most events are added to the SBML model as `sbml:reaction`. For instance, the text trigger and event annotation corresponding to E1 in Example 1 result in the following SBML description

```
<reaction metaid="metaid_0000E1"
sboTerm="SBO:0000216"
id="E1"
name="Phosphorylation"
reversible="false">
<annotation> ... </annotation>
</reaction>
```

The SBO/GO term is assigned according to the mapping depicted in Table 2. The reaction id is based on the event id (E10). The metaid of the form "metaid_0000 + id" is also added and the `sbml:reaction` name is the event-type. Lastly, all reactions are constructed as non reversible.

In a second step `sbml:reactant`, `sbml:product` and `sbml:modifier` are added to SBML reactions based on the roles of events.

`Theme` is the entity that undergoes the effects of the event. It is mapped to the `sbml:reactant` of the SBML reaction. For this a reactant reference is created and the species corresponding to the entity is linked to that reference via the id of the species (annotation-id of the entity).

`Product` can be specified for Binding, Dissociation[2] and Conversion events. Product is mapped to `sbml:product` of the corresponding reaction. The entities appearing in the product role are used for creating a product reference with the same entity.

[2] In data used for evaluation we also encountered Dissociation events with `Participant` and `Complex` roles. They are mapped to `sbml:product` and `sbml:reactant` respectively.

Standoff Event	SBO/GO term	SBO/GO name
Conversion	SBO:0000182	conversion
Acetylation	SBO:0000215	acetylation
Deacetylation	GO:0006476	Protein Deacetylation
Methylation	SBO:0000214	Methylation
Demethylation	GO:0006482	Protein Demethylation
Phosphorylation	SBO:0000216	phosphorylation
Dephosphorylation	SBO:0000330	Methylation
Ubiquitination	SBO:0000224	Ubiquitination
Deubiquitination	GO:0016579	Protein Deubiquitination
Degradation	SBO:0000179	degradation
Catabolism	GO:0009056	Catabolic Process
Catalysis	SBO:0000172	Catalysis
Protein_catabolism	GO:0009056	Catabolic Process
Association	SBO:0000177	non-covalent binding
Binding	SBO:0000177	non-covalent binding
Dissociation	SBO:0000180	dissociation
Regulation	GO:0065007	biological regulation
Positive_regulation	GO:0048518	positive regulation
Activation	SBO:0000412	biological activity
Negative_regulation	GO:0048519	negative regulation
Inactivation	SBO:0000412	biological activity
Gene_expression	GO:0010467	Genetic Production
Transcription	SBO:0000183	Transcription
Translation	SBO:0000184	Translation
Localization	GO:0051179	Localization
Transport	SBO:0000185	Transport Reaction
Pathway	SBO:0000375	Process

Table 2. Mapping of event annotations to SBO/GO term.

`Cause` is an entity/event causing the event. Cause is eventually mapped to entities which are then mapped to the reaction as `sbml:modifier` (via modifier reference).

Information in `Site` (which describes the site on the Theme entity that is modified in the event) is added to the "Notes" section of the SBML reaction as there seems to be no direct way to represent this information in SBML. Notes are human-readable annotations that can be added to SBML reactions.

Step 4: Handle Localization and Transport Events Localization and Transport events are handled differently from other events. They occur with additional roles besides `Theme`.

`AtLoc` describes the location/compartment at which the entity/species is located not an actual reaction. Hence, localization events with `AtLoc` roles do not end up as reactions in SBML. Instead, first we check if a `sbml:compartment` described by the `AtLoc` role exists, else a new `sbml:compartment` is created (see the nucleus example discussed earlier). Next, the compartment of the theme entity of the event is set to the corresponding `sbml:compartment`.

`FromLoc/ToLoc` Transport and Localization events can also include `FromLoc` and `ToLoc` roles which describes the transport of the theme entity/species from some location/compartment

to another. Consequently, we create a reaction where the Theme entity/species starts out in the compartment described by FromLoc (sbml:reactant) and ends up in the compartment described by the ToLoc (sbml:product) role. If the FromLoc/ToLoc sbml:compartment does not exist when creating the sbml:reaction, a new sbml:compartment is created corresponding to FromLoc/ToLoc.

Step 5: Handle Gene Expression Events We model Gene expression events (e.g. Transcription and Translation) as reactions in SBML. However, this class of reactions does not have the sbml:reactant role. For Transcription events (process in which a gene sequence is copied to produce RNA) if the type of Theme is RNA, it gets mapped to sbml:product. If the type of Theme is DNA, then it gets mapped to the sbml:modifier of the Transcription sbml:reaction. Translation events are handled in a similar manner.

Step 6: Handle Regulation Events In principle regulation events such as Positive/Negative Regulation, Activation and Inactivation can be handled as described in Step 3 when the Theme and Cause are species. If Theme and Cause are species then they are added to a regulation reaction as reactant and modifier respectively.

However, the standoff format definition also allows regulation events where Theme and Cause are themselves events[3]. For example, the following standoff lines describe a Positive regulation of a Phosphorylation event.

```
T14 Protein 776 782 eIF-4E
T15 Protein 852 859 insulin
T43 Phosphorylation 820 835 phosphorylation
T44 Positive_regulation 839 848
increased
E21 Phosphorylation:T43 Theme:T14
E22 Positive_regulation:T44 Cause:T15
Theme:E21
```

If the Theme is an event, then we do not create a reaction but simply add the Cause entity as a modifier to the reaction corresponding to the Theme event of the regulation. For the example above this means that the Phosphorylation reaction E21 is positively regulated (modified) by insulin (T15).

In reality though things are a bit more complicated since the Theme event might itself not exist as a reaction. For instance, there could be an event description as follows:

```
E23    Positive_regulation:T35 Cause:T21
Theme:E13
E13    Positive_regulation:T36 Theme:E21
```

Here, the event E23 has Theme E13, which itself is a Positive regulation with Theme E21. However, E13 itself does not correspond to a reaction. In this case the algorithm recursively tracks down the Theme event across multiple event annotations until it finds an event that exists in the SBML model as a reaction (In this case E21 is identified as the Theme for E23).

[3] In some of the data used to test our conversion we also encountered Catalysis events which had event themes. They are handled exactly as Positive Regulation events.

In case the Cause is an event, the product of the Cause event is used as a modifier. If the reaction corresponding to the Cause event does not have a product yet, then a corresponding product species is first created and added to the model.

Step 7: Optional Cleanup and Annotation Operations As a last step optional cleanup/enhancement operations can be performed. They can be used to ensure consistency of the resulting SBML model.

Add UniProt information We use the annotation-text to retrieve information about species from UniProt. The UniProt ID is added as controlled vocabulary to the corresponding SBML species. Other information is added as XML annotation and XHTML notes. This includes information about alternate names, gene names, gene ids where available and appropriate.

Remove unused species Not all entities end up as products, reactants or modifiers of an SBML reaction. In many cases, the named entity recognizer might recognize some entity but no links to events is established. However, the entities might have been added to the model (see Step 1). Entities not partaking in any reaction can be removed automatically.

Complete reactions The software supports automatic adding of products and reactants for reactions that were not explicitly annotated in that way. For instance, all phosphorylation events can extended with corresponding sbml:product species. The completion takes into account that certain reactions such as Gene expression reactions do not have reactants. Here is an example of this. For a Phosphorylation reaction, the first pass of the algorithm maps the Theme to sbml:reactant and no sbml:product is added. For example, E1 (in Example 1) would have Akt1 as a sbml:reactant. To complete this reaction a new sbml:species with name phoAkt1 is created representing the phosphorylated form of Akt. phoAkt1 is added as the sbml:product to the reaction E1 (See Figure 2).

Remove reactions without reactants, products In some cases the standoff file might include events that cannot be translated into reactions with reactants and/or products. For example, we encountered in real data that a reaction might only have a modifier (Cause). Such reactions are automatically removed if requested by the user.

1.2 Implementation

We used python and the python version of libSBML to develop the conversion algorithm. libSBML was used for generating and accessing the SBML model content. We used a custom implementation of a Standoff parser which translates the line-wise description of standoff triggers and events in a1/a2 and ann files into a memory structure of triggers (id, type, text) and events (id, type, roles). These structures are the basis for generating and completing the SBML model. The conversion is fast. It scales linearly with the number of entities, events and roles. The conversion algorithm is available from https://github.com/sbnlp.

Dataset	# species	# reactions	# edges
MTOR-HMN	2242	777	2457
MTOR-ANN	2457	857	2343
MTOR-NLP	292049	100130	203042

Table 3. Number of species, reactions and edges between them for the different datasets.

2 DATASETS

In addition to MTOR-HMN and MTOR-NLP we also used another dataset MTOR-ANN. The following gives a comprehensive overview.

MTOR-HMN is a mTOR pathway map manually constructed by human expert pathway curators (Caron *et al.*, 2010). The pathway is encoded in a dialect of SBML used by CellDesigner (Funahashi *et al.*, 2008). We convert the CellDesigner format into pure SBML and annotate reactions and species further by automatically assigning reaction types and gene/protein identifers (see Section **??**).

MTOR-ANN consists of 57 abstracts of scientific papers from Pubmed related to the mTORpathway map. The data set was *human-annotated* for NLP system training (Ohta *et al.*, 2011, Corpus annotations (c) GENIA Project[4]). This corpus gives an idea of the potential performance of a machine with human-level NLP extraction capabilities. Annotated NLP entities and events were used to create SBML representations and further annotated using various tools (discussed below).

MTOR-NLP consists of 522 full text papers mentioned in the mTOR pathway map. Paper pdfs were downloaded automatically and translated into raw txt files using CERMINE (Tkaczyk *et al.*, 2015). We managed to extract text from 501 papers. The 501 papers were processed using the Turku Event Extraction System mentioned earlier. From the extracted NLP events we created SBML representations of pathway maps for each text using (Spranger *et al.*, 2015). The SBML was further annotated using various tools (discussed below) and, finally, loaded into a single pathway map.

Role of different datasets MTOR-ANN and MTOR-NLP are different in how they are constructed and consequently what kind of conclusion we can draw from them. MTOR-ANN is a human-annotated dataset which contains much less data than MTOR-NLP. However, because it is human-annotated it allows us to evaluate a human-level performance extraction systems. We cannot expect that MTOR-ANN is able to reconstruct everything in MTOR-HMN (recall). However, it appears reasonable to expect that what is annotated in MTOR-ANN does occur in MTOR-NLP (high precision). In other words, MTOR-ANN allows us to test whether annotated data (ST) is in principle sufficient for solving the pathway curation problem.

Table 3 shows number of species, reactions and edges between them for the different datasets.

[4] http://nactem.ac.uk/GENIA/current/Other-corpora/mTOR-Pathway-Events/

Dataset	coverage	unique signatures	unique terms
MTOR-HMN	89.92%	443	538
MTOR-ANN	86.75%	207	320
MTOR-NLP	84.50%	6220	4194

Table 4. Entrez gene annotation results species.

Dataset	coverage	unique signatures	unique terms
MTOR-HMN	100%	36	16
MTOR-ANN	100%	13	13
MTOR-NLP	100%	9	9

Table 5. SBO annotation results reactions.

	MTOR-HMN	MTOR-ANN	MTOR-NLP
acetylation	1	0	0
activation	82	105	16485
association	210	211	21055
conversion	71	0	0
deacetylation	4	0	0
dephosphorylation	61	14	0
deubiquitination	18	0	0
dissociation	43	55	0
gene expression	4	46	18810
localization	0	16	474
negative regulation	61	101	10723
phosphorylation	172	252	25406
protein catabolism	24	18	1080
regulation	0	54	4832
transcription	78	8	1265
translation	23	1	0
transport	87	53	0
ubiquitination	20	4	0

Table 6. Reaction types extracted and annotated for various data sets. All reactions are annotated with their most specific type. Numbers are non-cumulative. For instance, the 171 conversion operations in MTOR-HMN are only annotated with the general conversion (SBO:182) and not more specific reaction types.

3 ANNOTATION RESULTS

As discussed in the main text, one of the crucial aspect of the comparisons we perform relies on the assumption that all the reactions and species can be identified even if, for instance, their names are not exactly the same. Therefore, we annotate species and reactions in all pathways with additional information

Species annotation results We annotate species using GNAT which for each species name returns a set of Entrez gene identifiers - we call these sets Entrez gene signatures. Table 4 shows that between 80 and 90% of species where annotated for each dataset. The table also shows that there are far more Entrez gene identifiers and unique signatures in MTOR-NLP than in MTOR-HMN. This demonstrates that there are far more species implicated in MTOR-NLP than there are in MTOR-HMN.

Reaction annotation results In terms of the reaction types, we achieve 100% coverage for all three datasets (see Table 5). MTOR-NLP has only 9 unique signatures compared to 36 in human and 13 in MTOR-ANN as shown in 5. In terms of specific reaction types, Table 6 shows the exact statistics. To note is that MTOR-NLP is unable to recognize specific reaction types. This might be due to limitations of the learning algorithm or training data.

4 MEASURING OVERLAP

We developed various *matching strategies* for species and reactions taking into account ways these different nodes are represented. In particular, we developed matching strategies with different constraints on when two species or two reactions are considered equal. Some of these matching strategies are very demanding. Others are looser. This is done on purpose to quantify the amount of overlap between species/reactions in MTOR-HMN and MTOR-ANN/MTOR-NLP in various ways. Obviously though, more loose matching strategies might make things equal which really should not be.

4.1 Species

Species can be equal if their names are equal but that might not be enough. Names can be synonymous and an exact string match might not reveal that. To overcome this, we annotated species with Entrez gene identifiers (a controlled vocabulary). In order, to decide whether a species in MTOR-HMN is the same as in MTOR-ANN and/or MTOR-NLP we use Entrez gene identifier matching. Identifiers of a species are basically so various matching strategies will make different assumptions about the constraints on sets of Entrez gene identifiers.

nmeq: Two species are equal if their names are exactly equal. We remove certain prefixes from the names (e.g. phosphorylated).

appeq: Two species are equal if their names are approximately equal. Two names are approximately equal iff their Levenshtein-based string distance is above 90 (Levenshtein, 1966)

enteq: Two species are equal if their entrez gene identifiers are exactly equal. This basically translates two the two species bqbiol:is identifier sets being exactly the same (order does not matter).

entov: Two species are equal if their entrez gene identifiers sets overlap. This basically translates to the two species bqbiol:is identifier sets overlapping.

wc: MTOR-HMN contains complex species but MTOR-NLP and MTOR-ANN do not. We therefore also allow species in MTOR-NLP and MTOR-ANN to match with constituents (*wc*) of complexes in MTOR-HMN. A link present in MTOR-ANN/MTOR-NLP between some protein and its phosphorylated version, will match if a link is present in a complex that contains that protein in MTOR-HMN. *wc* combines with all other species matching strategies. For instance *nmeq/wc* - means that two species are equal if their names match exactly or if one of them is a part of a complex and the name of the complex matches.

4.2 Reactions

Reaction types such as phosphorylation etc are encoded using a controlled vocabulary of SBO/GO terms. Reactions in MTOR-HMN sometimes have various reaction types associated with them. So for instance, there might be a reaction that is both a phosphorylation and an activation. In MTOR-ANN and MTOR-NLP, each reaction only has a *single* type. To account for this we developed various strategies for determining whether two reaction match.

sboeq: Two reactions are equal if their signatures are exactly the same. That is, the whole set of SBO/GO terms of one reaction is the same as of the other reaction.

sboov: Two reactions are equal if their signatures overlap. That is, the intersection of the set of SBO/GO terms of one reaction is with the set of SBO/GO terms of the other reaction is not empty.

sobisa: Two reactions are equal, iff requires that there is at least one SBO/GO term in each signature that relate in a is_a relationship in the SBO reaction type hierarchy. For instance, if there is a phosphorylation reaction and a conversion reaction, then *sboisa* will match because phosphorylation is a subclass of conversion according to the SBO type hierarchy.

4.3 Edges

We only allowed strict edge matching. So if an edge marks a reactant, then it has to be a reactant in MTOR-HMN. Same holds for product and modifier.

4.4 Quantifying Performance using F-score

We are interested in understanding the amount of overlap between the human curated target data (MTOR-HMN) and the automatic extraction data (MTOR-NLP). In essence this makes the problem an multi-class classification problem and we can use various measures for quantifying the amount of overlap between MTOR-HMN and MTOR-NLP. We use here *micro-averaged F-score*, precision and recall (Sokolova and Lapalme, 2009). Suppose we measure the overlap of the target graph t (MTOR-HMN) and source graph s (either MTOR-NLP or MTOR-ANN). F-score is defined as the harmonic mean of precision and recall. Precision measures the amount of false positives vs true positives in the source graph. Recall quantifies the amount of true positives vs false negatives in the target graph. Here are the formulas for with tp - true positives, fn - false negatives, fp - false positives and P - precision, R - recall, and F - f-score

$$
\begin{aligned}
P &= \frac{\mathrm{tp_s}}{\mathrm{tp_s} + \mathrm{fp_s}} \\
R &= \frac{\mathrm{tp_t}}{\mathrm{tp_t} + \mathrm{fn_t}} \\
F &= \frac{PR}{P + R}
\end{aligned}
$$

We use what is called a macro-averaged F-score measure, where all classes of reactions and species are counted in the same way even though there is much less data for certain reactions and species. We apply these measure in two ways

	MTOR-HMN	MTOR-ANN	MTOR-NLP
# species	2242	2567	292049
# names	582	367	27928
# appr names	568	324	21920
# Entrez signatures	443	207	6220

Table 7. Overview of number of species, unique names, approximate unique names and unique Entrez signatures.

4.4.1 Node matching: We measure the overlap of species and reactions separately. For instance, for each reaction in MTOR-NLP/MTOR-ANN we try to find matching reactions in MTOR-HMN. We allow for multiple overlaps, that is each reaction in MTOR-NLP/MTOR-ANN is allowed to match with multiple reactions in MTOR-HMN. We the count all overlapping reactions as true positive and all not matched reactions in MTOR-NLP/MTOR-ANN as false positives. All reactions in MTOR-HMN that did not match are counted as false negatives. Notice that the exact numbers for matching then depend on how strict matching is. For instance, a matching strategy such as sboeq will have less true positives (and consequently more false negatives and positives) then a matching strategy that is less strict, e.g. sboisa. The same process is applied for species (with different matching strategies). In node matching an F-Score of 100 means that all species or reactions in MTOR-HMN have been perfectly identified by species in MTOR-NLP/MTOR-ANN.

4.4.2 Subgraph matching: We also use the matching strategies to measure the overlap of subgraphs in MTOR-NLP and MTOR-ANN with MTOR-HMN. For this each subgraph in MTOR-NLP/MTOR-ANN is matched with MTOR-HMN. If there is a subgraph isomorphism for that graph given a particular combination of node matching strategies, then we count this as a true positive and all species, reactions and edges in the matched graph and subgraph are counted as true positives. For instance, a strategy such as *nmeq, sboeq* requires that species are equal by name, reactions are equal by sboeq standards and edge labels are exactly the same. In this scenario a F-score of 100 means that all nodes, reactions and edges present in MTOR-HMN are matched by a subgraph present in MTOR-NLP/MTOR-ANN and there are no graphs in MTOR-NLP/MTOR-ANN that did not match with anything in MTOR-HMN. Before measuring F-score, we remove all isolated nodes from MTOR-NLP and MTOR-ANN.

5 RESULTS: SPECIES

Table 7 shows detailed results of the number of unique species per dataset. MTOR-HMN contains 443 species with different Entrez gene signatures of 2242 total species. The numbers are even more sever for MTOR-NLP that contains 292,049 species but only 6330 unique signatures..

Table 8 gives results for species overlap. Here we reduce the number of species by names, approximate and Entrez gene signatures for each dataset and then measure the unique species overlap. Numbers are quite low except for *entov* matching strategy, which shows that MTOR-NLP recalls almost all of the species in MTOR-HMN. MTOR-ANN recalls almost half the species in MTOR-HMN. However precision is generally quite low compared to recall. The only outlier here is MTOR-HMN/MTOR-ANN in *entov* matching which has a very high precision.

	precision	recall	f-score
MTOR-HMN/MTOR-ANN			
nmeq	20.71	13.06	16.02
appeq	28.07	16.32	20.64
enteq	44.44	20.77	28.31
entov	83.57	55.76	66.89
MTOR-HMN/MTOR-NLP			
nmeq	0.96	45.88	1.87
appeq	1.60	51.55	3.10
enteq	12.66	98.70	22.44
entov	58.04	99.55	73.33

Table 8. Unique species retrieval score comparing various matching methods.

	MTOR-HMN/ MTOR-ANN			MTOR-HMN MTOR-NLP		
matchfn	prec	rec	f-score	prec	rec	f-score
nmeq	50.60	31.89	39.13	33.66	64.94	44.34
appeq	59.41	37.69	46.12	35.71	72.39	47.83
enteq	64.43	40.19	49.50	48.96	72.93	58.59
entov	81.03	58.97	68.26	67.15	89.74	76.82
nmeq/wc	50.60	41.97	45.89	33.66	78.19	47.06
appeq/wc	56.41	47.32	51.47	35.65	83.01	49.88
enteq/wc	81.03	61.15	69.70	67.15	91.30	77.38
entov/wc	81.03	61.15	69.70	67.15	91.30	77.38

Table 9. Species overlap.

	MTOR-HMN	MTOR-ANN	MTOR-NLP
# reactions	777	2420	100130
# SBO/GO signatures	36	13	9
# SBO/GO terms	16	13	9

Table 10. Overview of number of reactions, unique SBO/GO signatures and SBO/GO terms.

Table 9 measure the total species overlap for various datasets. Results show rather high recall high for matching strategies such as *entov* and expecially *entov/wc*.

6 RESULTS: REACTIONS

Table 11 details the number of reaction, unique SBO/GO signatures and terms for all datasets. MTOR-HMN contains 777 reactions with 12 SBO/GO terms, i.e. reaction types. MTOR-ANN contains 12 and MTOR-NLP slightly less. Each reaction can have multiple SBO/GO terms associated with it. We call this the SBO/GO signature of a reaction. For instance, a particular reaction can be typed as phosphorylation and activation. Its signature are then the SBO/GO terms for these 2 reactions. The table shows that this actually only

	precision	recall	f-score
MTOR-HMN/MTOR-ANN			
sboeq	69.23	25.00	36.73
sboov	84.62	88.89	86.70
sboisa	92.31	91.67	91.99
MTOR-HMN/MTOR-NLP			
sboeq	55.56	13.89	22.22
sboov	77.78	72.22	74.90
sboisa	88.89	80.56	84.52

Table 11. Unique reactions retrieval score for different SBO matching strategies.

	MTOR-HMN/ MTOR-ANN			MTOR-HMN MTOR-NLP		
matchfn	prec	rec	f-score	prec	rec	f-score
sboeq	84.86	69.63	76.49	74.84	56.76	64.55
sboov	92.09	85.97	88.93	94.70	72.59	82.18
sboisa	98.19	95.11	96.63	99.53	84.68	91.51

Table 12. Reactions overlap. These are results for all reactions (not just unique reactions).

happens in MTOR-HMN. Human annotators are free to combine various reactions into a single reaction if they see fit. There is no replication of this in the automated data.

Table 12 shows how much reactions overlap across the different datasets. MTOR-ANN catches 1/4 of the reaction SBO/GO signatures directly and up to 91% when we allow for overlap sbo_is_a relationship. MTOR-NLP only directly includes 1 out of 8 reaction signatures. However, the overlap is higher when allowing for reaction SBO/GO signatures to overlap and individual SBO terms to be in a is_a relationship. These results also show that there are reactions in MTOR-NLP and MTOR-ANN that are not part of MTOR-HMN.

7 RESULTS: NETWORK CONNECTEDNESS

Ultimately we are interested in networks of reactions and species. Studying the output of NLP systems it becomes immediately clear that the result of these systems differs from hand-curated data in an important aspect: *connectedness*. To show this we measured isolation of species and networks (reactions cannot be isolated for structural reasons in SBML).

Table 14 shows the number of subgraphs and isolated nodes in all datasets. In MTOR-HMN there are 4 separate subgraphs (no connection between them) and 19 isolated nodes. All of these except for one big subgraph are modeling mistakes by human curators. MTOR-HMN is one connected network. On the other hand, MTOR-ANN and MTOR-NLP consist of numerous unconnected networks. Each of them is quite small as the data in Table 14 demonstrates by showing min, max, mean and median number of species and reactions (nodes) in each connected component subgraph. Results show that subgraphs in MTOR-ANN and MTOR-NLP on average contain between 2 and 3 species and reactions. For the most part MTOR-NLP and MTOR-ANN consist of scattered small connected regions of reactions and species.

Dataset	# subgraphs	# isolated nodes
MTOR-HMN	4	15
MTOR-ANN	490	662
MTOR-NLP	83093	110490

Table 13. Number of isolated subgraphs (no connection between them) and isolated nodes (single node subgraphs).

Dataset	min	mean	median	max
MTOR-HMN nodes	6	751	18	2962
MTOR-HMN edges	5	912	19	3606
MTOR-ANN nodes	2	6	4	24
MTOR-ANN edges	1	5	4	28
MTOR-NLP nodes	2	3	3	215
MTOR-NLP edges	1	2	2	465

Table 14. Statistics of nodes and edge distributions for subgraphs

8 RESULTS: NETWORK OVERLAP

Table 15 (Figures 3 and 4) shows precision, recall, F-score for max overlap of different matching strategies. The table shows results for MTOR-ANN and MTOR-NLP successively. In general the first rows (*nmeq, sboeq*) represent very strict matching strategies. The last row (*appeq/entov/wc, sbois*) shows results for more "relaxed" strategies and include complex constituent matching.

Let us first analyze the performance of MTOR-NLP. The automated NLP system is able to recall 10% of all edges given the strictest matching strategy. This means that 1 in 10 edges in MTOR-HMN appear in MTOR-NLP. Also, if we look at the most relaxed matching strategy *appeq/entov/wc, sbois*, we find that roughly 2 of 3 edges and 1 of 2 nodes (species and reactions) in the human curated MTOR have something to do with the NLP extracted data.

Matching strategies that allow for matching complex constituents always have a higher recall and precision performance than their non constituent matching counterparts. For instance, *nmeq, sboeq* matches almost 36 percentage points less edges than *nmeq/wc, sboeq* (MTOR-HMN/MTOR-NLP). This increase in performance of constituent matching points to the fact that human modelers often attribute reactions to the whole complex. For instance, a phosphorylation may be acting on a constituent of a complex but the human modeler chooses to connect the reaction with the whole complex. These matching strategies do account for that and therefore are able to improve the numbers (in some cases) considerably.

Reactions in MTOR-HMN are sometimes incorporating various reaction types. In MTOR-ANN and MTOR-NLP, on the other hand, each reaction only has a single type. Reaction matching strategies *sboov* and *sboisa* account for that by looking at overlaps. This means that reactions in MTOR-ANN and MTOR-NLP will match with a reaction MTOR-HMN if the reaction type signatures intersection is not empty. In reality this means that the reaction in MTOR-ANN or MTOR-NLP has to be an element of the reaction in MTOR-HMN.

Lastly, let us take a look at MTOR-HMN/MTOR-ANN. MTOR-ANN contains much less data than MTOR-NLP but the reason we include it here is because MTOR-ANN consists of human annotated data. It therefore gives an idea about the limits of the annotation data and the limits of human annotation. If all of

the problems discussed so far are purely a problem of the NLP system, then MTOR-ANN should do better than MTOR-NLP in terms of precision but not in terms of recall. Recall will be low because the MTOR-ANN consists of less data. However, we would expect high precision numbers. Interestingly, data shows that even for NLP-ANN precision is low. With relaxed matching strategies *appeq/enteq/wc, sboisa* and *appeq/entov/wc, sboisa*, we see some substantial recall 20% (remember NLP-ANN is only abstracts). Nevertheless precision for edges is only 12% and 15% for nodes.

REFERENCES

Caron, E., Ghosh, S., Matsuoka, Y., *et al.* (2010). A comprehensive map of the mtor signaling network. *Molecular systems biology*, **6**(1).

Funahashi, A., Matsuoka, Y., Jouraku, A., *et al.* (2008). Celldesigner 3.5: a versatile modeling tool for biochemical networks. *Proceedings of the IEEE*, **96**.

Hucka, M., Finney, A., Sauro, H., *et al.* (2003). The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**(4), 524–531.

Junker, B. H., Klukas, C., and Schreiber, F. (2006). Vanted: a system for advanced data analysis and visualization in the context of biological networks. *BMC bioinformatics*, **7**(1), 109.

Kim, J.-D., Pyysalo, S., Ohta, T., Bossy, R., Nguyen, N., and Tsujii, J. (2011). Overview of bionlp shared task 2011. In *Proceedings of the BioNLP Shared Task 2011 Workshop*, pages 1–6. Association for Computational Linguistics.

Le Novere, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., *et al.* (2009). The systems biology graphical notation. *Nature biotechnology*, **27**(8), 735–741.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710.

Nédellec, C., Bossy, R., Kim, J.-D., *et al.* (2013). Overview of bionlp shared task 2013. *ACL*, page 1.

Novere, N. L., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., Mendes, P., *et al.* (2005). Minimum information requested in the annotation of biochemical models (miriam). *Nature biotechnology*, **23**(12), 1509–1515.

Ohta, T., Pyysalo, S., and Tsujii, J. (2011). From pathways to biomolecular events: opportunities and challenges. In *Proceedings of BioNLP 2011 Workshop*, pages 105–113. ACL.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, **45**(4), 427–437.

Spranger, M., Palaniappan, S., and Ghosh, S. (2015). Extracting biological pathway models from nlp event representations. In *Proceedings of the 2015 Workshop on Biomedical Natural Language Processing (BioNLP 2015)*, pages 42—51. ACL.

Tkaczyk, D., Szostek, P., Fedoryszak, M., *et al.* (2015). Cermine: automatic extraction of structured metadata from scientific literature. *International Journal on Document Analysis and Recognition (IJDAR)*, **18**(4), 317–335.

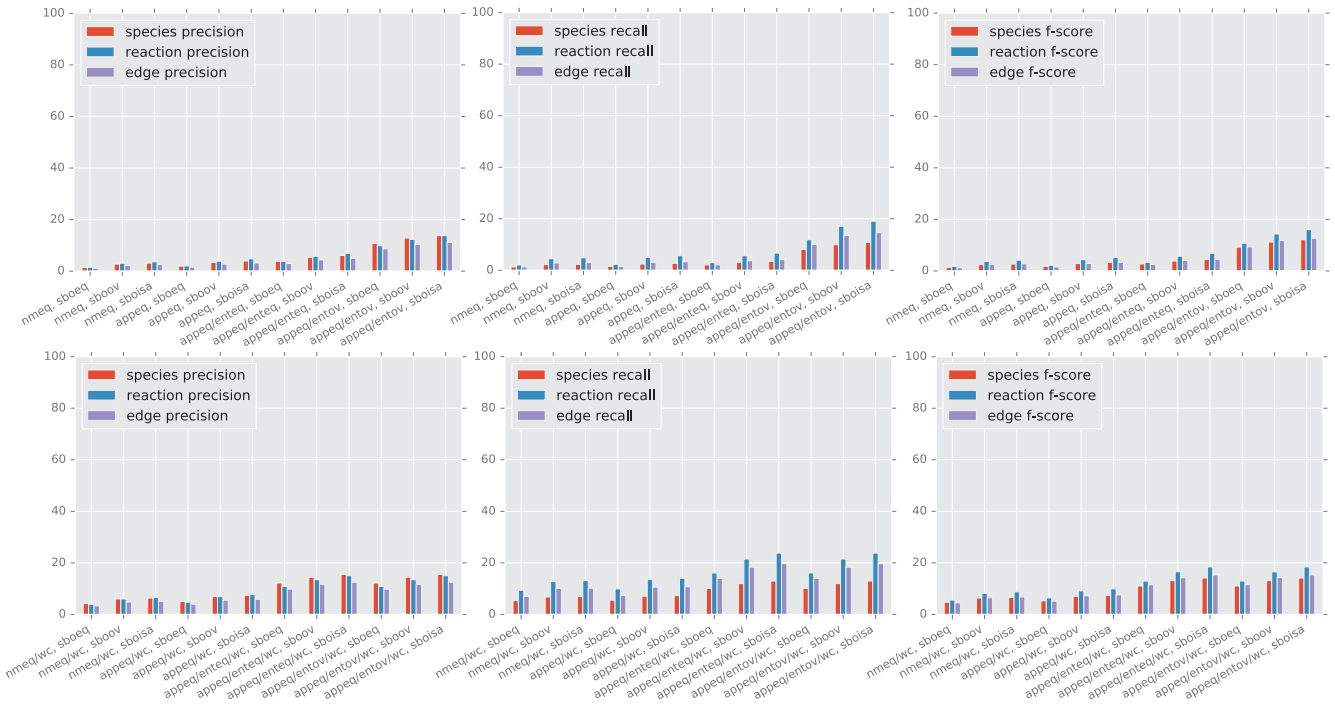

Fig. 3. MTOR-HMN/MTOR-ANN: Precision, recall, F-score (columns) for various matching strategies with or without constituent matching (rows) quantifying the overlap of MTOR-ANN with MTOR-HMN (data from Table 15 top rows).

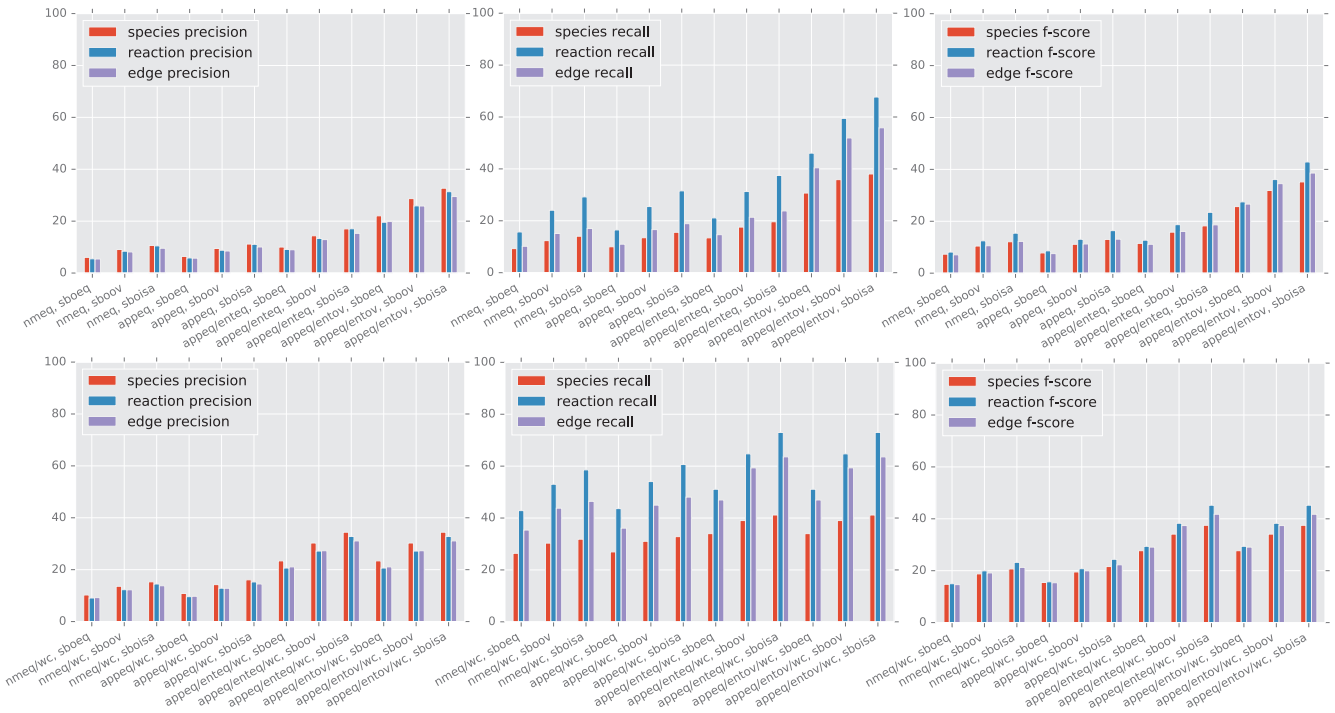

Fig. 4. MTOR-HMN/MTOR-NLP: Precision, recall, F-score (columns) for various matching strategies with or without constituent matching (rows) quantifying the overlap of MTOR-NLP with MTOR-HMN (data from Table 15 bottom rows).

MTOR-HMN/MTOR-ANN												
	nodes			species			reactions			edges		
	prec	rec	f-score	prec	rec	f-score	prec	rec	f-score	prec	rec	f-score
nmeq, sboeq	1.43	1.52	1.48	1.42	1.34	1.38	1.47	2.06	1.71	1.12	1.34	1.22
nmeq, sboov	2.83	2.78	2.81	2.73	2.19	2.43	3.05	4.50	3.64	2.23	2.85	2.50
nmeq, sboisa	3.26	2.98	3.12	3.10	2.32	2.65	3.62	4.89	4.16	2.52	2.97	2.73
appeq, sboeq	1.97	1.72	1.84	1.94	1.52	1.70	2.03	2.32	2.17	1.53	1.51	1.52
appeq, sboov	3.51	3.11	3.30	3.36	2.45	2.84	3.84	5.02	4.35	2.73	3.09	2.90
appeq, sboisa	4.19	3.48	3.80	3.94	2.72	3.22	4.75	5.66	5.16	3.18	3.34	3.26
appeq/enteq, sboeq	3.76	2.32	2.87	3.78	2.10	2.70	3.73	2.96	3.30	2.98	2.16	2.50
appeq/enteq, sboov	5.48	3.68	4.40	5.35	2.99	3.84	5.76	5.66	5.71	4.34	3.74	4.02
appeq/enteq, sboisa	6.34	4.27	5.11	6.09	3.43	4.39	6.89	6.69	6.79	4.92	4.15	4.50
appeq/entov, sboeq	10.47	9.04	9.70	10.76	8.07	9.23	9.83	11.84	10.74	8.68	10.05	9.31
appeq/entov, sboov	12.69	11.76	12.21	12.86	9.90	11.19	12.32	17.12	14.33	10.41	13.55	11.78
appeq/entov, sboisa	13.76	12.92	13.33	13.75	10.79	12.10	13.79	19.05	15.99	11.12	14.61	12.63
nmeq/wc, sboeq	4.12	6.39	5.01	4.20	5.35	4.71	3.95	9.40	5.57	3.31	6.96	4.48
nmeq/wc, sboov	5.99	8.31	6.96	5.98	6.78	6.36	5.99	12.74	8.15	4.79	10.05	6.49
nmeq/wc, sboisa	6.42	8.51	7.32	6.35	6.91	6.62	6.55	13.13	8.74	5.08	10.18	6.78
appeq/wc, sboeq	4.95	6.66	5.68	5.04	5.53	5.27	4.75	9.91	6.42	3.97	7.37	5.16
appeq/wc, sboov	6.88	8.74	7.70	6.88	7.09	6.98	6.89	13.51	9.13	5.50	10.58	7.23
appeq/wc, sboisa	7.46	9.01	8.16	7.35	7.27	7.31	7.68	14.03	9.93	5.87	10.74	7.59
appeq/enteq/wc, sboeq	11.76	11.63	11.69	12.18	10.08	11.03	10.85	16.09	12.96	9.79	13.96	11.51
appeq/enteq/wc, sboov	14.12	14.41	14.26	14.44	11.95	13.08	13.45	21.49	16.54	11.65	18.32	14.24
appeq/enteq/wc, sboisa	15.34	15.77	15.55	15.49	12.98	14.12	15.03	23.81	18.43	12.48	19.70	15.28
appeq/entov/wc, sboeq	11.76	11.63	11.69	12.18	10.08	11.03	10.85	16.09	12.96	9.79	13.96	11.51
appeq/entov/wc, sboov	14.12	14.41	14.26	14.44	11.95	13.08	13.45	21.49	16.54	11.65	18.32	14.24
appeq/entov/wc, sboisa	15.34	15.77	15.55	15.49	12.98	14.12	15.03	23.81	18.43	12.48	19.70	15.28

MTOR-HMN/MTOR-NLP												
	nodes			species			reactions			edges		
	prec	rec	f-score	prec	rec	f-score	prec	rec	f-score	prec	rec	f-score
nmeq, sboeq	5.84	10.93	7.62	6.04	9.28	7.32	5.49	15.70	8.13	5.41	10.18	7.07
nmeq, sboov	8.83	15.37	11.22	9.07	12.36	10.46	8.41	24.07	12.46	8.13	15.10	10.57
nmeq, sboisa	10.58	17.95	13.31	10.64	14.05	12.11	10.47	29.21	15.41	9.56	17.01	12.24
appeq, sboeq	6.19	11.66	8.09	6.40	9.99	7.80	5.82	16.47	8.60	5.74	10.99	7.54
appeq, sboov	9.23	16.56	11.85	9.47	13.47	11.12	8.79	25.48	13.07	8.50	16.61	11.24
appeq, sboisa	11.14	19.68	14.22	11.17	15.57	13.01	11.07	31.53	16.39	10.04	18.88	13.11
appeq/enteq, sboeq	9.69	15.40	11.90	10.02	13.43	11.47	9.09	21.11	12.71	8.98	14.65	11.13
appeq/enteq, sboov	13.99	21.10	16.83	14.36	17.57	15.80	13.33	31.27	18.69	12.89	21.37	16.08
appeq/enteq, sboisa	17.04	24.21	20.00	17.02	19.63	18.23	17.06	37.45	23.44	15.30	23.81	18.63
appeq/entov, sboeq	21.17	34.65	26.28	22.05	30.69	25.66	19.57	46.07	27.47	19.84	40.42	26.62
appeq/entov, sboov	27.69	41.93	33.35	28.68	35.86	31.87	25.89	59.46	36.07	25.84	51.93	34.51
appeq/entov, sboisa	32.22	45.68	37.79	32.69	38.05	35.17	31.36	67.70	42.86	29.48	55.84	38.58
nmeq/wc, sboeq	9.83	30.64	14.89	10.25	26.40	14.77	9.08	42.86	14.98	9.23	35.37	14.64
nmeq/wc, sboov	13.13	36.17	19.26	13.59	30.33	18.77	12.28	53.02	19.95	12.24	43.79	19.14
nmeq/wc, sboisa	15.02	38.72	21.64	15.31	31.85	20.68	14.48	58.56	23.21	13.80	46.44	21.28
appeq/wc, sboeq	10.39	31.24	15.60	10.84	26.94	15.46	9.59	43.63	15.73	9.76	36.10	15.36
appeq/wc, sboov	13.74	36.97	20.04	14.24	31.04	19.52	12.85	54.05	20.77	12.82	44.97	19.95
appeq/wc, sboisa	15.80	39.98	22.64	16.09	32.83	21.59	15.27	60.62	24.39	14.50	48.07	22.28
appeq/enteq/wc, sboeq	22.41	38.39	28.30	23.39	33.99	27.71	20.63	51.09	29.39	21.07	46.97	29.09
appeq/enteq/wc, sboov	29.17	45.64	35.60	30.28	39.03	34.10	27.17	64.74	38.28	27.30	59.34	37.39
appeq/enteq/wc, sboisa	33.84	49.35	40.15	34.43	41.17	37.50	32.78	72.97	45.24	31.08	63.57	41.75
appeq/entov/wc, sboeq	22.41	38.39	28.30	23.39	33.99	27.71	20.63	51.09	29.39	21.07	46.97	29.09
appeq/entov/wc, sboov	29.17	45.64	35.60	30.28	39.03	34.10	27.17	64.74	38.28	27.30	59.34	37.39
appeq/entov/wc, sboisa	33.84	49.35	40.15	34.43	41.17	37.50	32.78	72.97	45.24	31.08	63.57	41.75

Table 15. Overlap MTOR-ANN and MTOR-NLP with MTOR-HMN.