Creation of a Suitable SNePS Representation Scheme for Medical Discharge Reports

Michael Schiller
CSE 663: Advanced Knowledge Representation
April 30, 2002

Abstract:

This is a continuation of an investigation into the translation of medical discharge reports into everyday understandable English. This will be done using a parser to translate the original reports into a SNePS representation, and then a generator which uses the representation of the original report as well as appropriate SNePS-represented background knowledge to generate easily understood translations of the original reports.

This paper deals primarily with the representation in SNePS of the original discharge report. A standard set of SNePS case frames is decided upon for parsing, and a representation of one discharge report is attempted using those case frames. This is done with one eye on automation of the original parse and the other eye on ease of generation of the final translated report. There is also some discussion of the background knowledge that will need to be represented in order to produce an effective translation.

I. Introduction

When one goes to the doctor's office, it is easy for one to become overwhelmed by the vast amounts of seemingly impenetrable medical language that is commonly used by those working there. One of the places where this is most evident is in the discharge reports that follow a hospital stay. For doctors, these reports are a concise, yet informative summary of what happened over the course of the patient's stay. For patients, however, it is easy to get lost in the language that doctors commonly use, as well as the proper technical names of symptoms, tests, and medications. What we are attempting to do through this project is bridge this apparent language gap.

Last semester, Frank Lehoullier, Rajeev Sood, and myself each continued work on this project by each proposing schemes for the representation of these discharge reports, and then providing actual representations of a report using those schemes. The result of that work was proof that these discharge reports could accurately be represented in Shapiro's Semantic Network Processing System (SNePS). Unfortunately, the eventual result was three different representation schemes that had little to do with each other.

Over the course of this paper, I attempt to rectify those disparities by proposing a standard set of case frames, which could theoretically be used to represent *any* discharge report. Taken into consideration will be the ease with which sentences can be parsed using the proposed case frames, and the ease of generation of a suitable plain-English translation from a representation using those case frames. I will also outline ideas for the representation of background knowledge using those case frames, and conclude with some explanation of the next steps that need to be taken in the overall project.

II. Case Frames Used for Representation

In coming up with a standard set of case frames, there were two distinct sets of case frames that I felt I needed to use. The first set of case frames are ones that deal with literal translation of the language used in the discharge report. This set consists largely of case frames from the SNePS Case Frame Dictionary, as well as previously written papers on representation of some certain ideas in SNePS. The second set of case frames are custom case frames that are specific to medical discharge reports, and are used specifically to identify those aspects of the discharge reports that most need translating of some sort.

Non-Discharge Report-Specific Case Frames:

Standard Case Frames:
- Lex
- Object / Propername
- Member / Class
- Object / Property
- Agent / Act
- Action / Object / Ind-Object
- Mod / Head
- Before / After
- Time
- Min / Max / Arg
- Object1 / Rel / Object2...ObjectN
- Object / Rel / Possessor

General Custom Case Frames:
- Object / Property / Value
- Object / Become / Time
- Years / Months / Days
- Hours / Minutes / Seconds
- Month / Day / Year

All of the standard case frames presented here has syntax and semantics that are outlined either in the case frame dictionary or previously published papers. The general custom case frames have functions that lend themselves to the representation of the

discharge reports. The Object-Property-Value case frame is used when a certain object has a property with a value associated with it. For example, it could be used for age, since age is a property of a living thing, and age also has a value associated with it. Object-Become-Time is used when the state of some object is changed over a certain period of time. This case frame is most handy as the health of a patient changes over the course of a hospital visit. Finally, Years-Months-Days and Hours-Minutes-Seconds are specialized case frames that literally represent spans of time, and Month-Day-Year represents a date. More detailed syntax and semantics are located in Appendix B.

Discharge Report-Specific Custom Case Frames:

Patient / Symptom / BodyPart
Patient / Medication
Patient / Diagnosis
Patient / Prescription / Duration
Patient / Test / Bodypart / Details (where "details" becomes a test-specific case frame)
    Airtype / Percent
    Count
    Reveal / Where
Sentence

These case frames were formulated with an eye on translation from the SNePS representation of the discharge report to English. The five case frames that center on "Patient" are what I have chosen as the most important things that need to be translated, all of which are prone to having complex medical names. Patient-Symptom-Bodypart details symptoms that the patient presents to the hospital. Patient-Medication represents medications administered to the patient during the patient's stay in the hospital. Patient-Diagnosis is the doctor's diagnosis of what is wrong with the patient, if anything. Patient-Prescription-Duration details any prescriptions given to the patient. Finally, Patient-Test-Bodypart-Details gives the representation of any tests performed on the

patient during the hospital visit. The “Details” arc points to another case frame that is specific to whatever test has been performed, thus giving every test its own unique case frame, to best represent the results of whatever test was given. In this case we have a white blood cell count, an X-ray, and a pulse ox, so appropriate case frames have been listed.

Finally, there is the Sentence case frame, which consists of an extra arc coming out of each asserted node in the original representation. At the end of the arc is simply a number indicating the sentence number that the asserted node originally appeared in. This is one possible way of preserving the narrative of the original discharge report; that is, the specific order of events. When those things get entered into a SNePS database, all sense of narrative is lost, as there’s no immediate way to pull the order things happened from the database. With the Sentence arc, all of the first sentence’s asserted nodes can be translated, then all of the second sentence’s, and so on, so that the translated discharge report can preserve the order of events found in the original.

The syntax and semantics of all of these case frames is found in Appendix B.

III. Representation of One Discharge Report

The previously described set of case frames was arrived upon by looking at a number of actual discharge reports, all centering around respiratory distress of some sort. As such, perhaps the best way I could think of to test the set I came up with was to actually represent one of the discharge reports with those case frames. I chose Teron’s discharge report at random (which is reprinted in its entirety as Appendix A), and came up with the representation illustrated in Appendix C.

The first decision I had to make was exactly which parts of the discharge reports I was going to try and represent. In the representations that had been done prior to this one, the amount of representation done ranged from “only the pertinent information” to “absolutely everything on the page.” My decision was that useful information could be found everywhere except in the last paragraph. I wanted to include the header at the top of each discharge report because information can be derived from that such as the name of the patient and the length of the hospital stay. The last paragraph, on the other hand, is the same for every discharge report, so at most, we merely need to come up with a canned paragraph of translated text for that paragraph rather than going through the trouble of representation.

With that out of the way, I found that there were a number of stumbling blocks on my way to developing an accurate representation. For example, the “Details” arc of Patient-Test-Bodypart-Details didn’t always exist. My original implementation was to have an individual case frame for every different possible test that could be run, as was my solution last semester. However, it was pointed out to me that it would likely be much more efficient in both translation phases to have one consistent “test” case frame that had one arc that pointed to details specific to whatever test was run. This would allow the language generator to identify tests more easily, and then just look at a database of “test” case frames to translate the details arc, rather than having to look through what could amount to over 100 case frames for each test.

When looking at the representation, one may also notice the absence of a couple of my custom discharge report-specific case frames. For example, I never use the Patient-Medication case frame. This is because I felt that the Patient-Medication case

frame didn't capture the spirit of the sentence in which I used it. So, instead of actually using the Patient-Medication case frame in this case, I used Agent-Act to represent that "the patient received the medication." Eventually, my intent is to use rules in the background knowledge to generate the custom case frames from the initial representation. More details on this can be found in section IV of this paper.

Finally, many of the sentences are written in the passive voice, which presents the problem of there being no apparent subject to the sentence. For a phrase such as "A home nebulizer machine was arranged for the patient," one's first instinct might be to make the home nebulizer machine the agent, "arranged" the action, and the patient the object. That interpretation loses the spirit of the Agent-Act case frame, however, as it's not actually the machine doing the arranging—rather, it's some unnamed medical worker that arranges the nebulizer machine for the patient. So, I create a SNePS build node, and make that node a member of class "medical worker," thus making the action "arranged," the object the home nebulizer machine, and the indirect object the patient. How exactly a parser will pick up that medical personnel perform these actions that are written in the passive, I don't know; that would be a topic for another paper, perhaps!

IV. Background Knowledge

Since I spent much of the semester working on perfecting the SNePS representation of the discharge report, I did not get a chance to actually represent any of the necessary background knowledge that will need to be stored for proper translation. I did, however, get a chance to formulate exactly what sorts of background knowledge will need to be represented. For the sake of this project, I divided the background knowledge up into two categories: rules and general knowledge. Rules create new case frames that

will be useful in the language generation stage. All rules will create custom case frames that fall into one of the five categories of things that need to be translated. General knowledge, on the other hand, expands on things in the knowledge base, and contains information on tests, medications, and anything else that the average person might not know about given only the name of what's being represented.

Necessary Rules for This Representation:

If a patient is begun on a medication of some sort, then add a Patient / Medication arc.
If a patient receives a medication of some sort, then add a Patient / Medication arc.
If a patient presents with something, then add a Patient / Symptom arc.

These rules solve the conflict between maintaining the spirit of the original discharge report and using the custom case frames I have provided. With these rules, the original representation can use case frames like Agent-Act to say that a patient is "begun" on something or a patient "receives" something, with the assumption that when the rules are run, if a patient "receives" or "is begun on" something, then there will be a Patient-Medication arc added to the knowledge base. In these cases, a "sentence" arc will probably need to be added to the Patient-Medication case frame with the same number as the activator for the rule. That is, if "patient was begun on…" has a sentence arc pointing to 3, the resulting Patient-Medication arc should also have a sentence arc pointing to 3. Similar rules can be followed with "patient presents with…" leading to a Patient-Symptom arc.

Teron's discharge report lent itself rather easily to the other three types of custom case frames, so no other rules are needed for this particular representation. It should be relatively straightforward, however, to see how similar rules could be created should similar problems arise in other discharge reports.

Possible General Background Knowledge for This Representation:

A patient is a person.
Zithromax is a medication.
A nurse is a medical worker.
A doctor is a medical worker.

A home nebulizer machine is a machine.
A home nebulizer machine is used at home.
A home nebulizer machine is a nebulizer.
A nebulizer turns liquid into mist.
A nebulizer gets the mist into the lungs of the person using it.
A nebulizer is generally used by children who are too small to use an inhaler.

Prelone is a brand of prednisolone.
Prednisolone is a medication.
Prednisolone is a steroid.
Prednisolone reduces swelling.
Prelone is a syrup.

Some of this background knowledge leads right to the rules—there's no way for the knowledge base to know that Zithromax is a medication unless it's specifically represented as such. The "received-to-Patient-Medication" rule won't work, however, unless it is known that Zithromax is indeed a medication. So, lists of all possible medications, symptoms, and so on need to be explicitly stated as such.

The rest of the background knowledge represented here is knowledge that would allow someone to understand better the tests, medications, symptoms, and so on listed in the discharge reports. Someone reading Teron's discharge report may want to know exactly what Prelone is and what it does. Using the background knowledge displayed above, the final translated discharge report can contain information as to what Prelone is, what it does, and even how it is administered. Similar information will be available for any other symptoms, medications, tests, and diagnoses. This is also the type of

information that we could obtain by contacting a medical professional and asking questions about whatever we don't already know in reference to these things.

V. Conclusions and Future Work

On first glance, it seems that I have come up with an effective set of case frames with which to represent discharge reports. However, more discharge reports need to be represented with the same set of case frames in order to insure that any situation can be represented gracefully with this particular set. Even over the course of my own representation, the set of case frames I was using changed dramatically—there's no reason to think that won't happen while using the same set on a different discharge report.

As work on this project continues, a parser and a language generator need to be programmed. My intent was such that the work I did this semester made parsing and generating as easy as those two tasks could possibly be. The sentence arc and the custom discharge report-specific case frames should make language generation much easier, and having a fixed set of case frames should make parsing possible, if nothing else.

Finally, the background knowledge that I have started to outline above should be represented in a fixed SNePS network, that can be accessed whenever translation needs to take place. In order to create such a database, it would be wise to contact a medical professional to get information on as many different medications, conditions, symptoms, and tests as possible. Using that information, we could create a translation scheme that makes even the most complex medical terms understandable, simply by explaining them right in the translated discharge report. Ultimately, I hope that my work on this project might be used as a launch pad for future work in the field of medical natural language processing.

Appendix A: Teron's Discharge Report

Patient name: Teron A
data birth 10/9/98
date of admission 9/17/00
date of discharge 9/18/00
no PMD

This is a 23 month old male with no significant past medical history who presented to the Children's Hospital emergency department with respiratory distress. In the emergency department, the patient received albuterol and Atrovent aerosols. A pulse ox was 86 percent on room air at presentation. A chest x-ray revealed a small right middle lobe opacity. Patient was afebrile and a white blood cell count was 10.9. The patient was begun on Prelone and Zithromax and admitted to the floor.

Overnight the patient improved markedly and by the morning had no oxygen requirement. Patient was tolerating medications well and albuterol aerosols were rapidly spaced. Patient was discharged to home with a diagnosis of RAD and was to continue albuterol aerosols and to complete a five day course of Prelone and Zithromax. A home nebulizer machine was arranged for the patient. Mom was asked to follow-up at Hodge pediatrics five days after discharge.

Thank you very much for this referral. A full discharge summary may be obtained from Children's Hospital by calling 716-878-7776. Please do not hesitate to contact us with any questions.

Sincerely,

Peter Winkelstein M.D.

Appendix B: Case Frames Used

The syntax and semantics for the following case frames can be found in “A Dictionary of SNePS Case Frames” (Shapiro, Rapaport, et al, 1994):

Lex
Object / Propername
Member / Class
Object / Property
Agent / Act
Action / Object / Ind-Object
Min / Max / Arg
Rel / Object1 / ObjectN

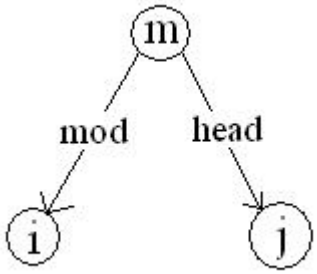
The syntax and semantics for the Time and Before / After / Duration case frames can be found in “Time In Narratives” (Almeida, 1995).

The syntax and semantics for the Object / Rel / Possessor case frame can be found in “SNePS Implementation of Possessive Phrases” (Chun, 1987).

The syntax and semantics for the rest of the case frames used in this paper can be found on the following pages.

Mod / Head

Syntax



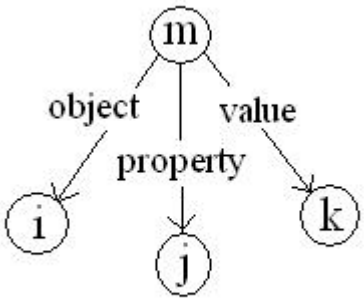
If i and j are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that head [[j]] is modified by modifier [[i]].

Object / Property / Value

Syntax



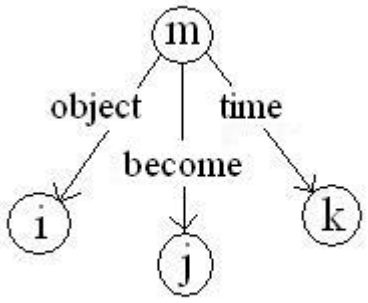
If i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that [[i]] has property [[j]] with value [[k]].

Object / Become / Time

Syntax



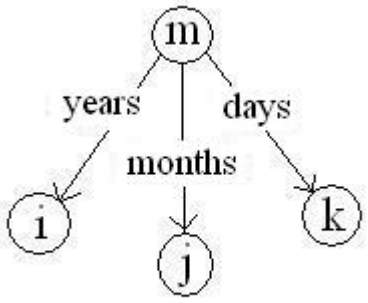
If i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that object [[i]] becomes [[j]] over time period [[k]].

Years / Months / Days

Syntax



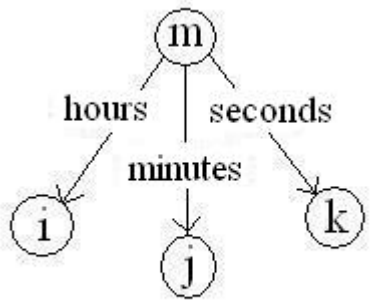
If i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured node.

Semantics

[[m]] represents the time period of [[i]] years, [[j]] months, and [[k]] days.

Hours / Minutes / Seconds

Syntax



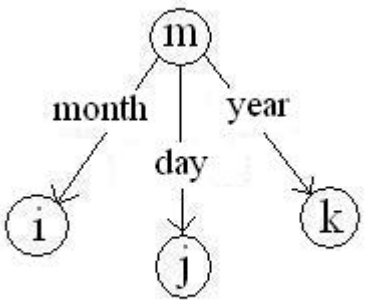
If i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured node.

Semantics

[[m]] represents the time period of [[i]] years, [[j]] months, and [[k]] days.

Month / Day / Year

Syntax



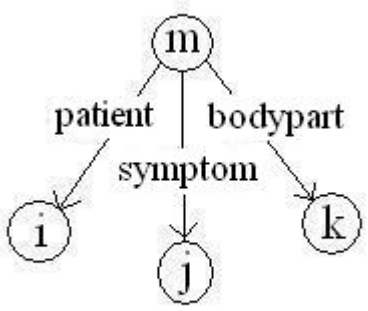
If i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured node.

Semantics

[[m]] represents the date of [[i]] / [[j]] / [[k]].

Patient / Symptom / Bodypart

Syntax



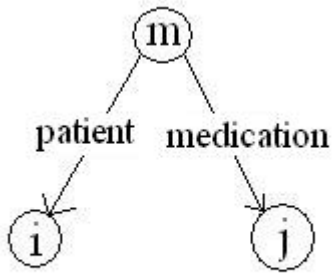
If i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that patient [[i]] went to the hospital with symptom [[j]] on part of his or her body [[k]].

Patient / Symptom

Syntax



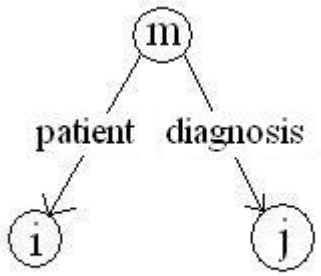
If i, and j are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that patient [[i]] took medication [[j]] while in the hospital's care.

Patient / Diagnosis

Syntax



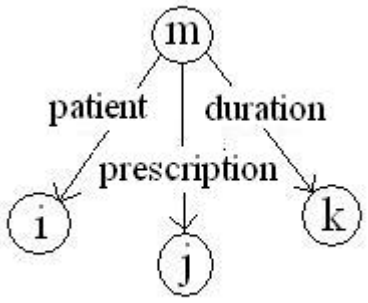
If i, and j are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that patient [[i]] was diagnosed with condition [[j]] over the course of the hospital stay.

Patient / Prescription / Duration

Syntax



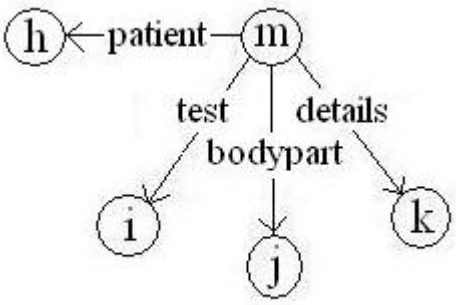
If i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that patient [[i]] was prescribed medication [[j]] for a duration of [[k]] time as a result of the hospital visit.

Patient / Test / Bodypart / Details

Syntax



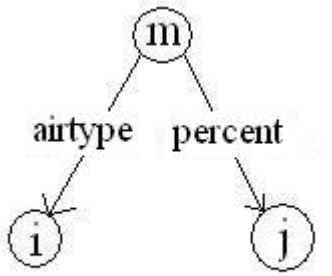
If h, i, j, and k are individual nodes and m is an identifier not previously used, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that patient [[h]] underwent test [[i]] on body part [[j]], which resulted in details [[k]].

Airtype / Percent

Syntax



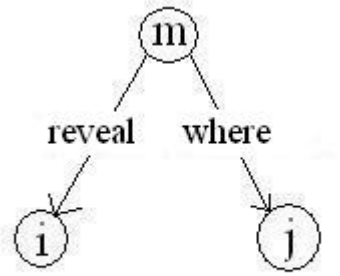
If i and j are individual nodes and m is an identifier pointed to by the details arc in a Patient / Test / Bodypart / Details case frame where the Test points to a pulse ox, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that the pulse ox was [[j]] percent on [[i]] type of air.

Reveal / Where

Syntax



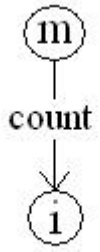
If i and j are individual nodes and m is an identifier pointed to by the details arc in a Patient / Test / Bodypart / Details case frame where the Test points to an X-ray, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that the X-ray revealed [[i]] in location [[j]] on the patient's body.

Count

Syntax



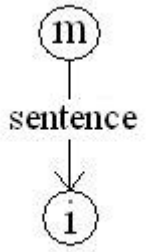
If i is an individual node and m is an identifier pointed to by the details arc in a Patient / Test / Bodypart / Details case frame where the Test points to a white blood cell count, then the above is a network and m is a structured proposition node.

Semantics

[[m]] is the proposition that the patient’s white blood cell count was [[i]].

Sentence

Syntax



If i is an individual node and m is a previously existing asserted structured proposition node, then the above is a network.

Semantics

[[m]] is a proposition that appeared in the original discharge report in sentence [[i]].

Appendix C: Representation of Teron's Discharge Report

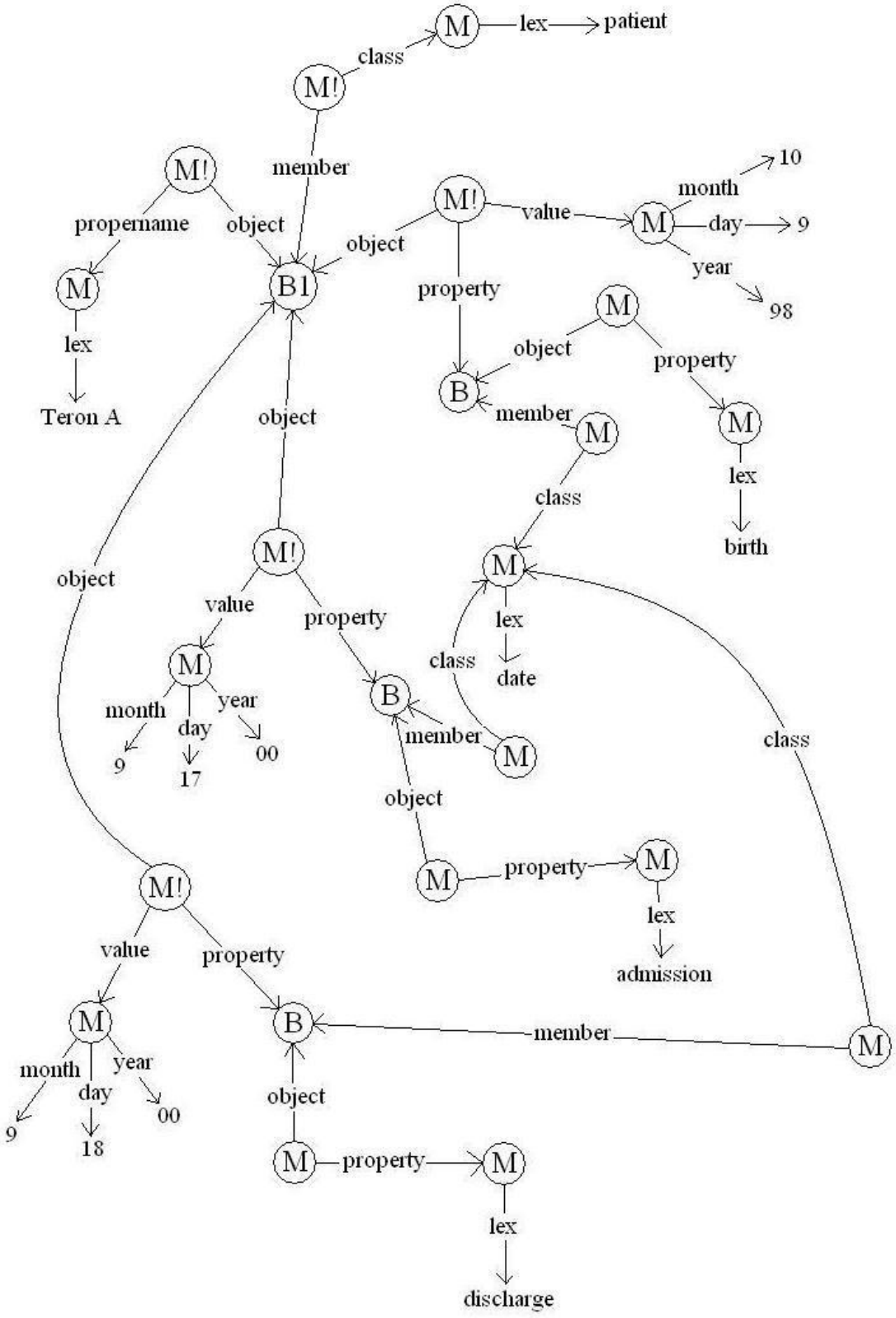
Notes on the Representation:

- The sentence arcs are not included in this representation, as they would clutter up the representation. Assume that a sentence arc with the number of the sentence would be added to each asserted node in the representation.

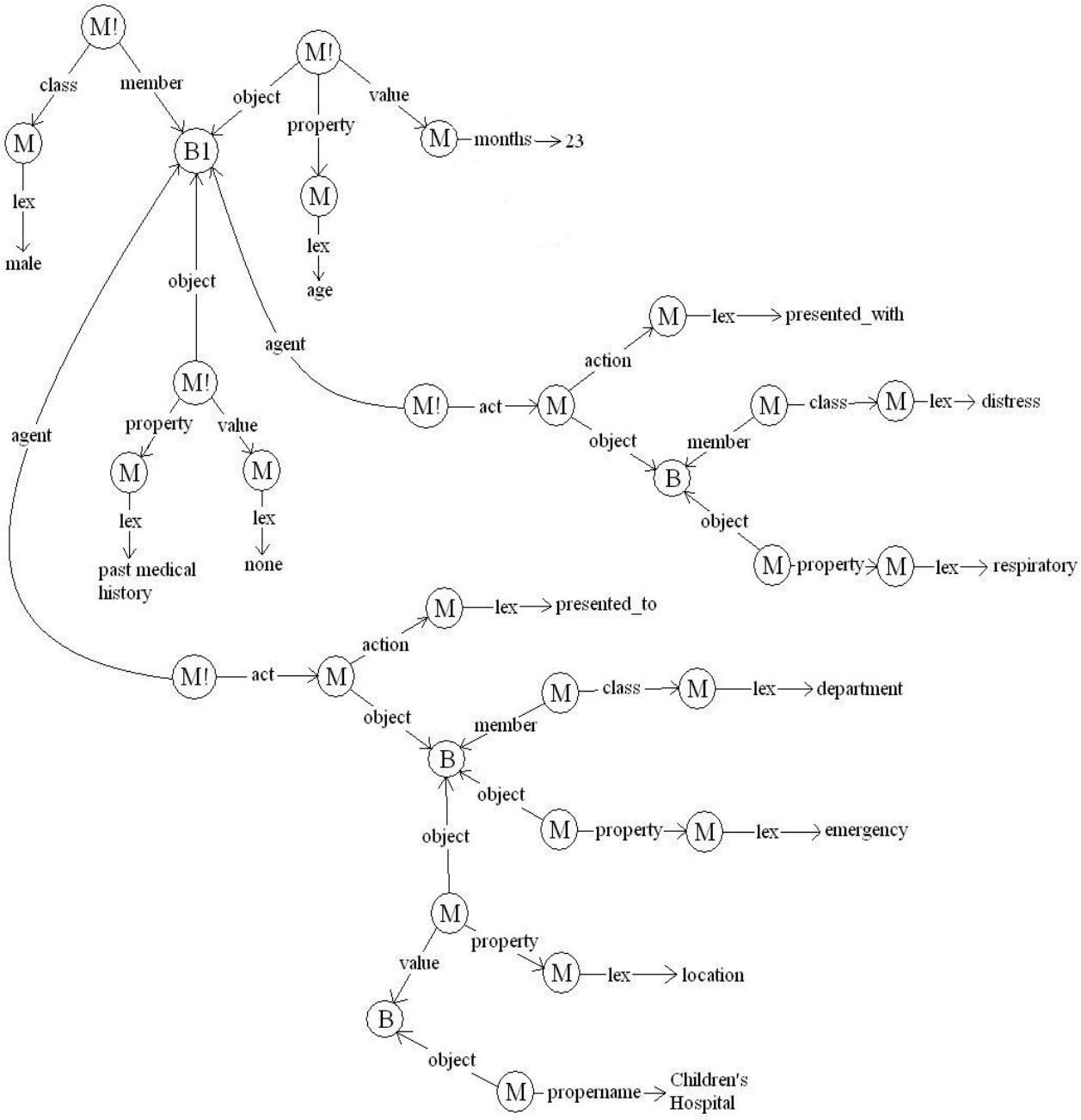
- If there is a proposition node with a proposition in parentheses next to it, it is assumed that the proposition represented by that node was previously introduced, and the arc pointing to that node would be pointing to that previously introduced node.

- Build node B1 can be found in the representation of each sentence—B1 is the Build node representing the patient.

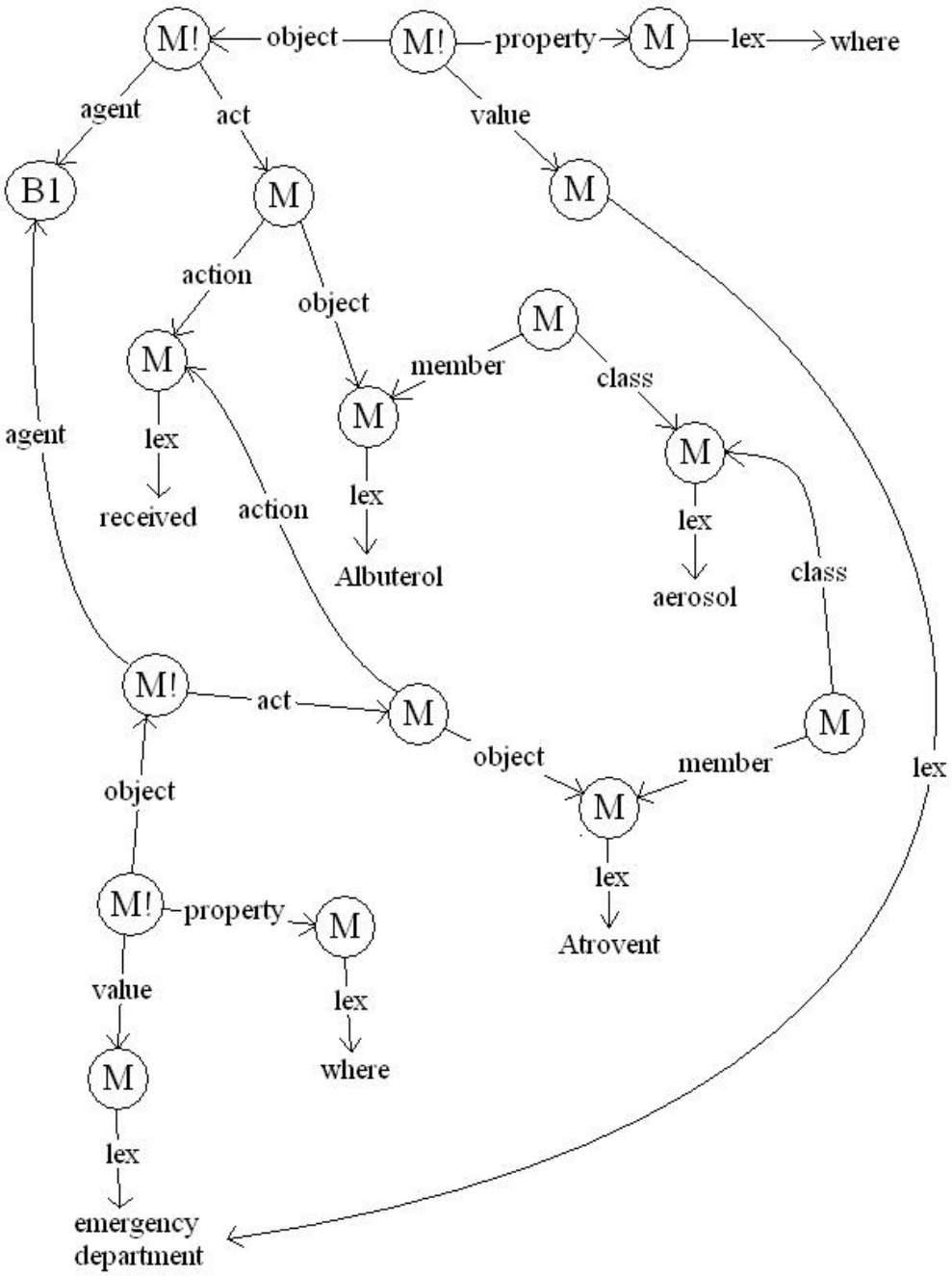
1. Patient name, Date of Birth, etc.:


M
lex
patient
class
M!
member
M!
propername
object
M
lex
Teron A
B1
object
M!
value
M
month
10
day
9
year
98
object
property
B
M
object
property
member
M
M
lex
birth
class
M!
value
property
M
lex
date
class
M
month
day
year
9
17
00
B
member
M
object
class
M
property
M
lex
admission
M!
value
property
M
month
day
year
9
18
00
B
member
M
object
M
property
M
lex
discharge
object

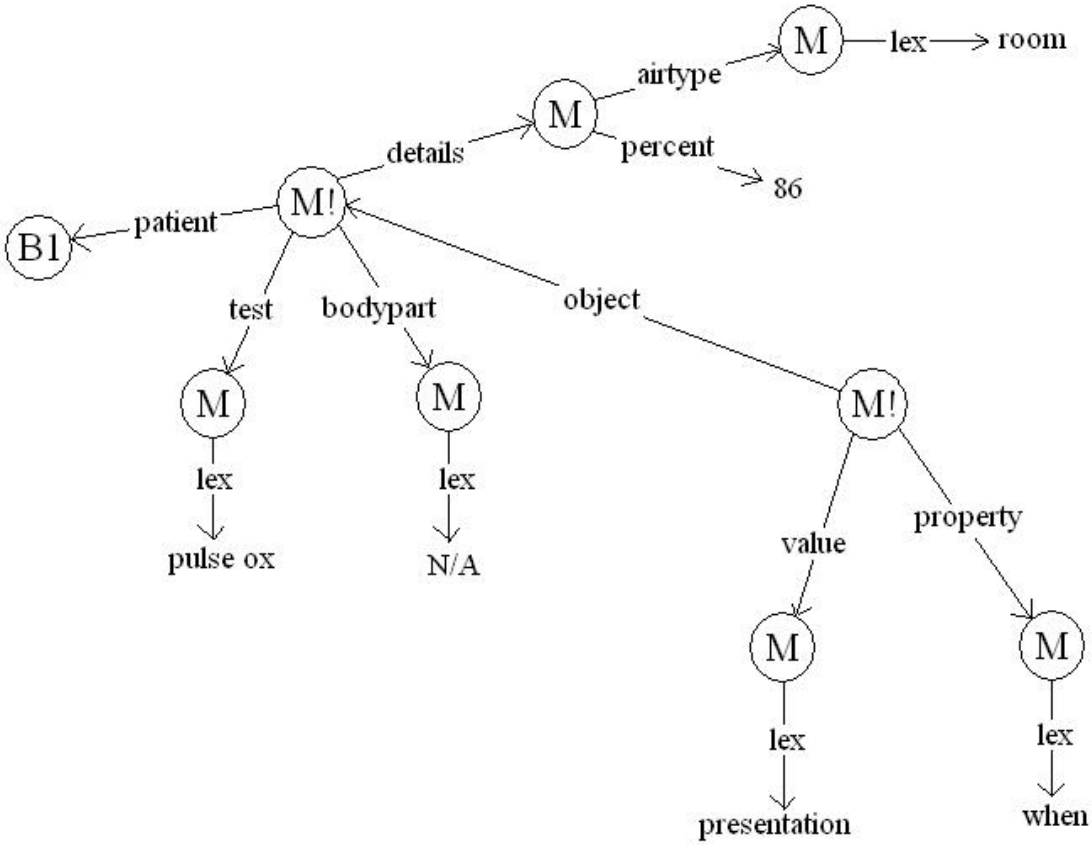
2. This is a 23 month old male with no significant past medical history who presented to the Children's Hospital emergency department with respiratory distress.



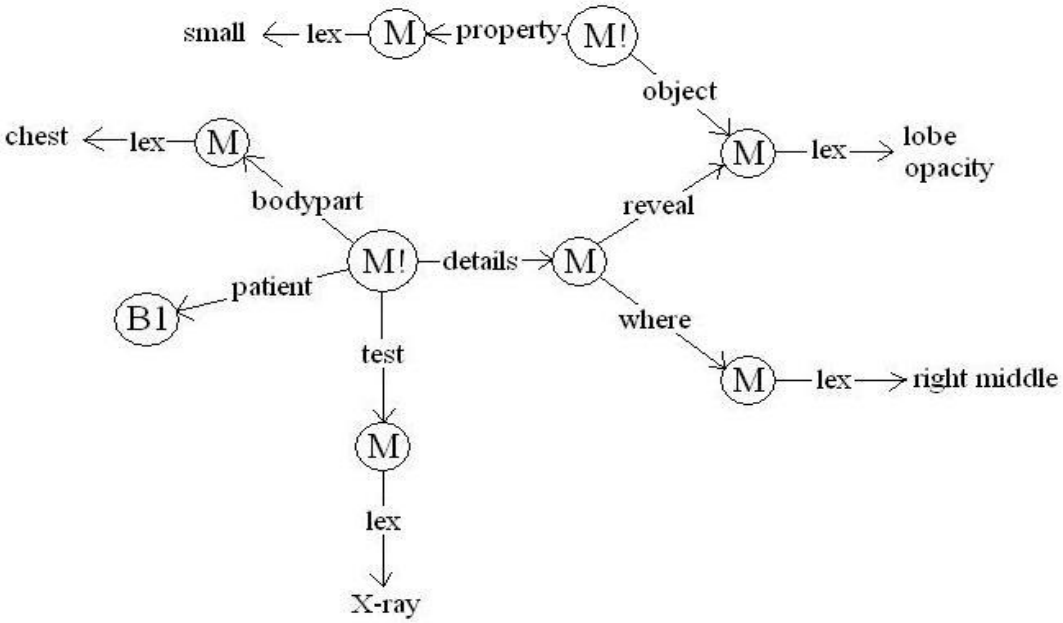
3. In the emergency department, the patient received albuterol and Atrovent aerosols.



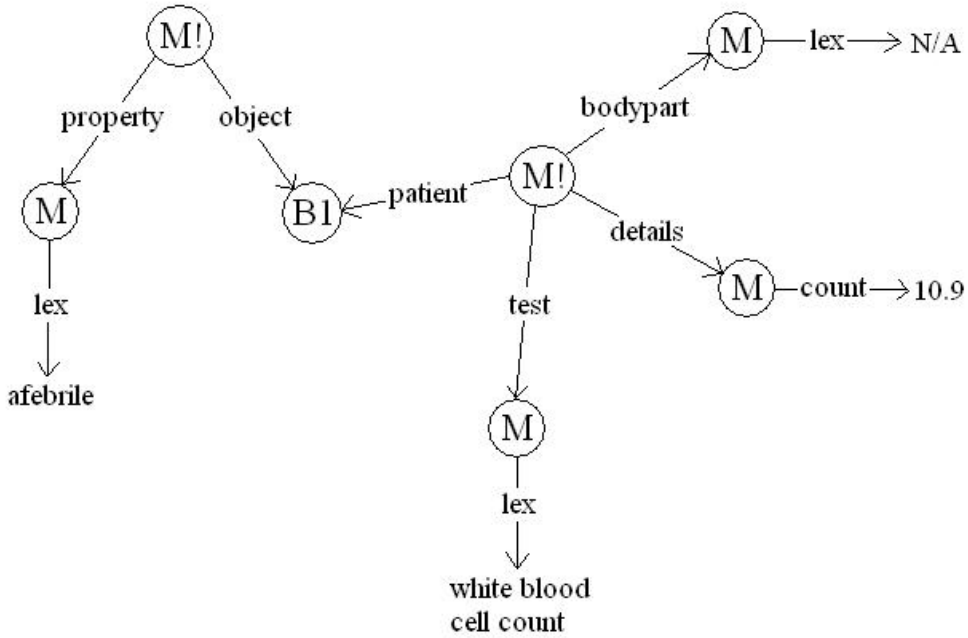
4. A pulse ox was 86 percent on room air at presentation.



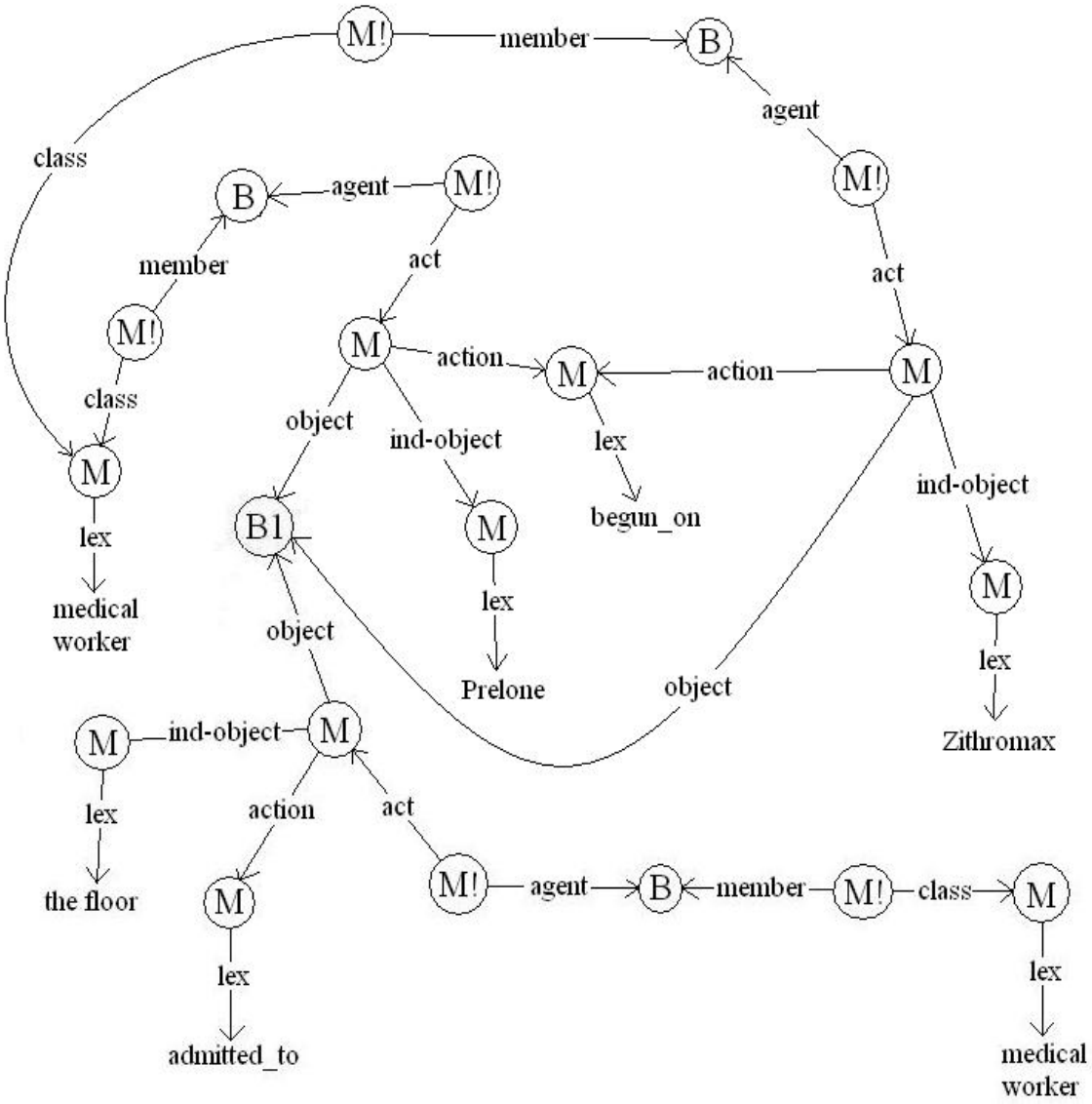
5. A chest x-ray revealed a small right middle lobe opacity.



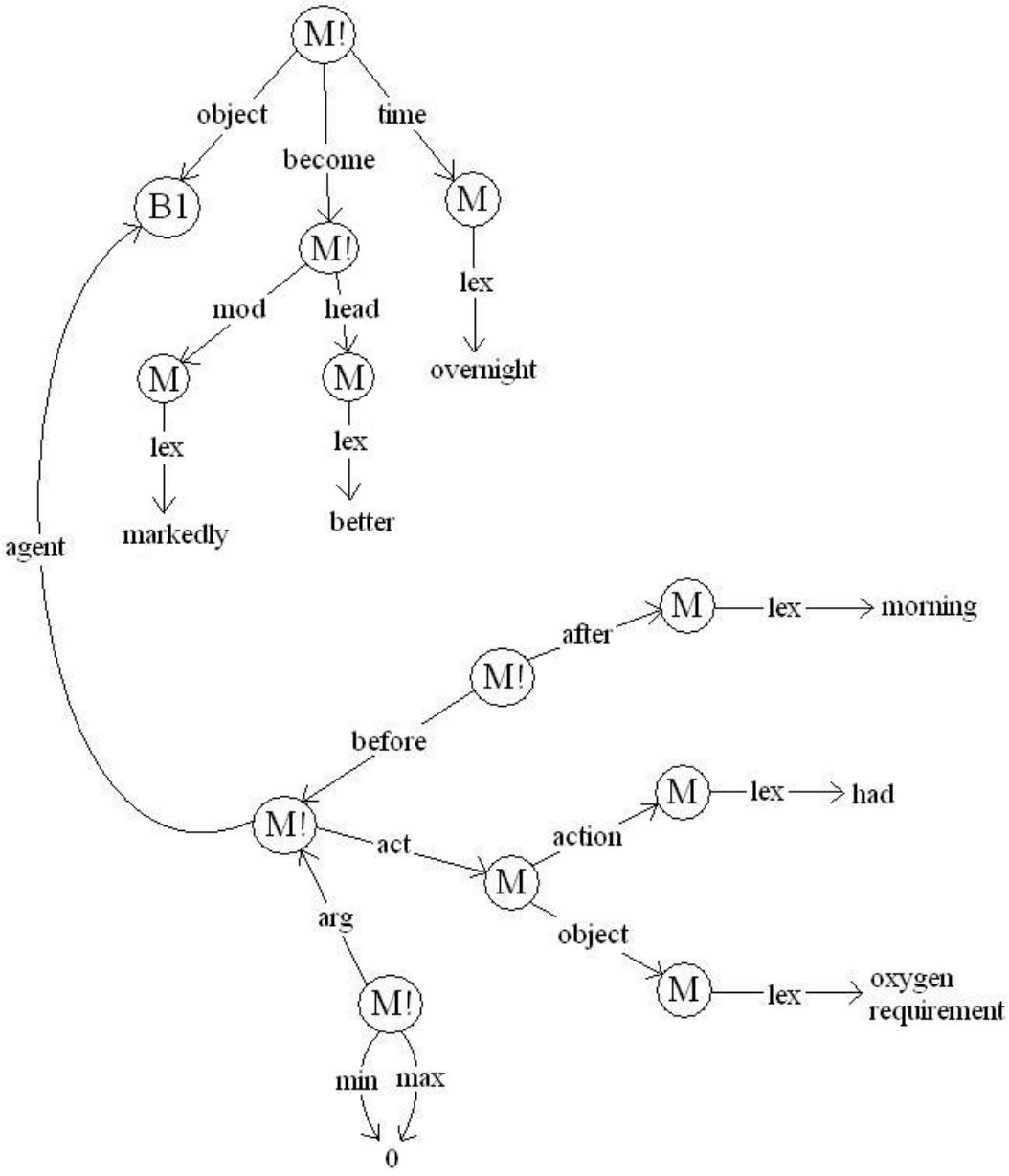
6. Patient was afebrile and a white blood cell count was 10.9.



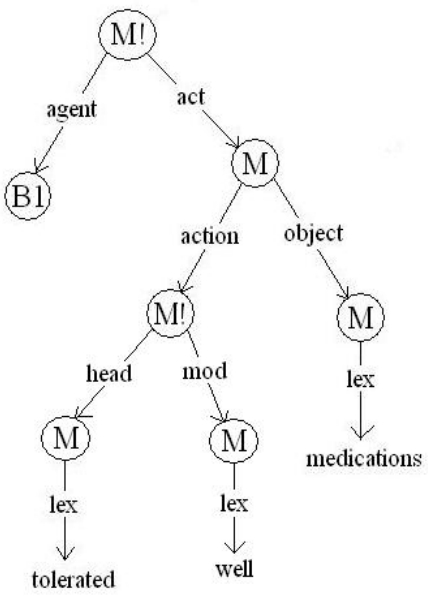
7. The patient was begun on Prelone and Zithromax and admitted to the floor.

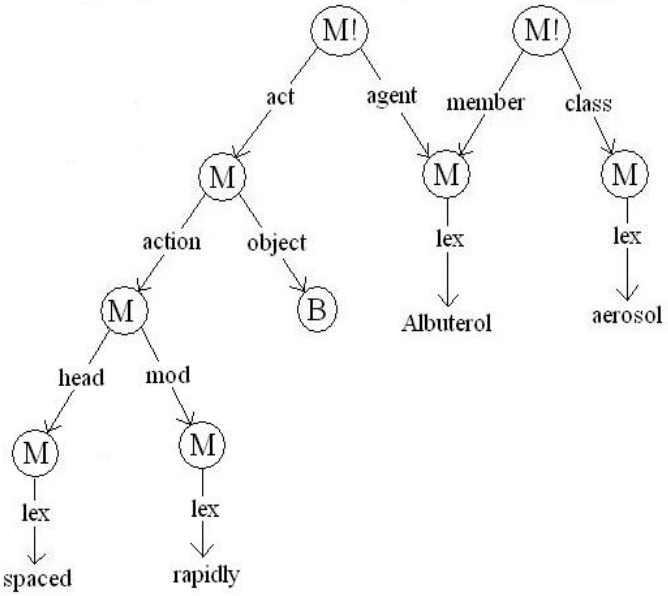


8. Overnight the patient improved markedly and by the morning had no oxygen requirement.

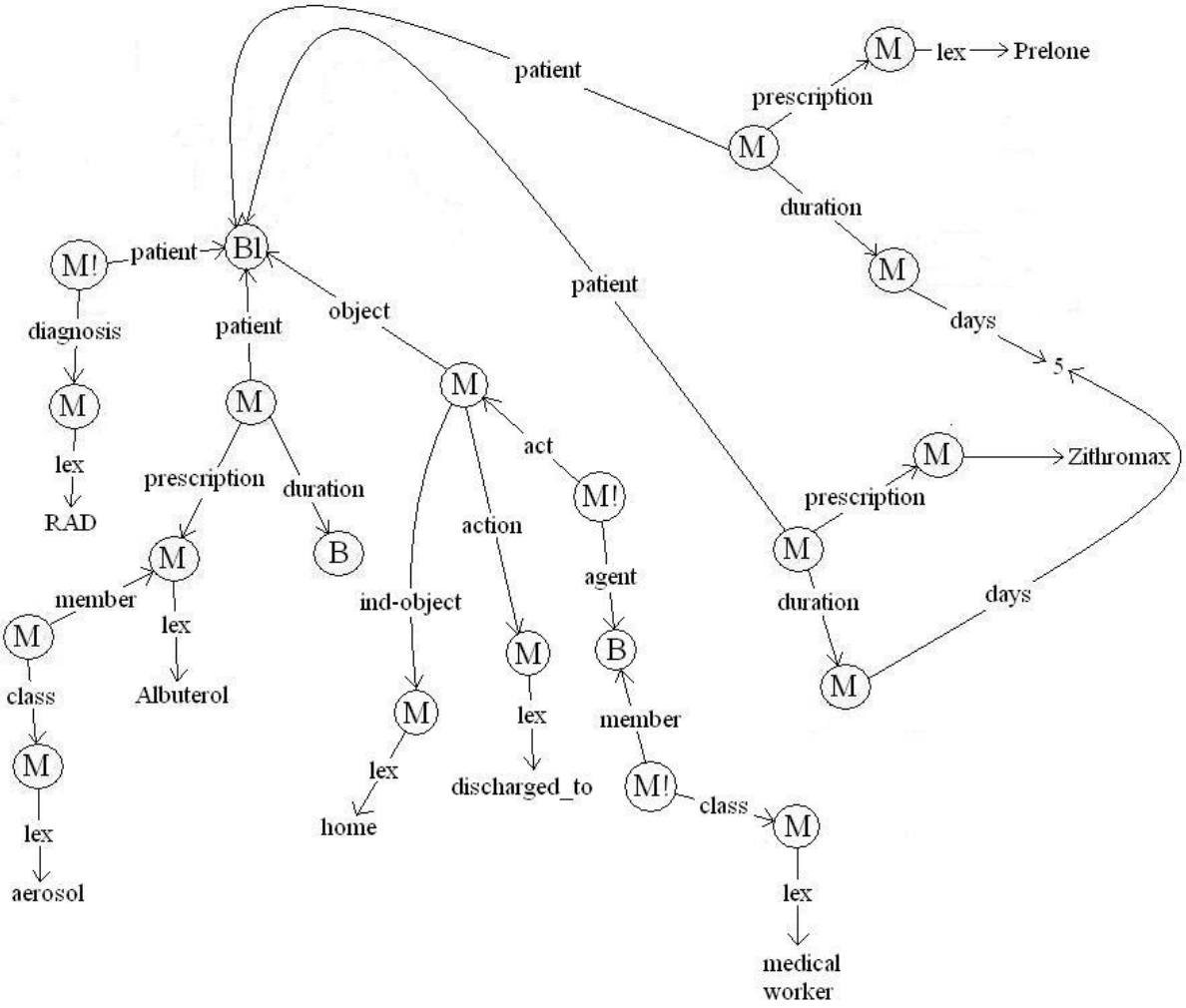


9. Patient was tolerating medications well and albuterol aerosols were rapidly spaced.

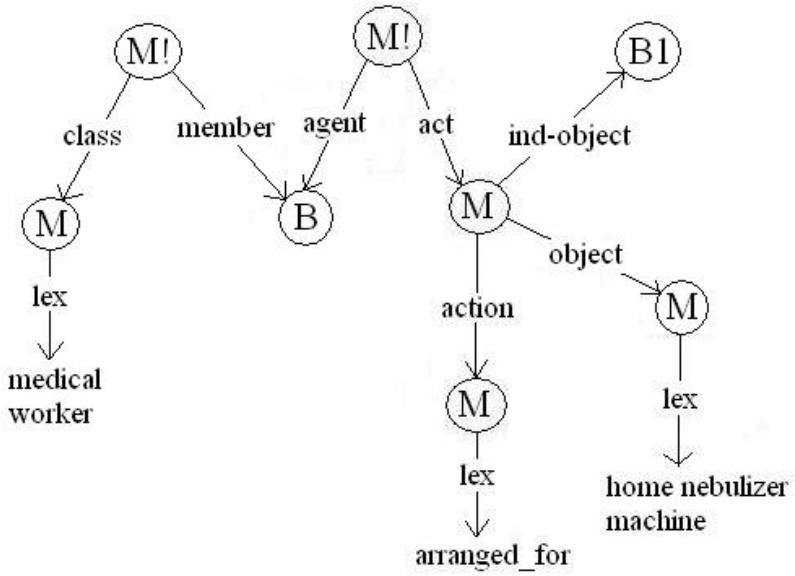




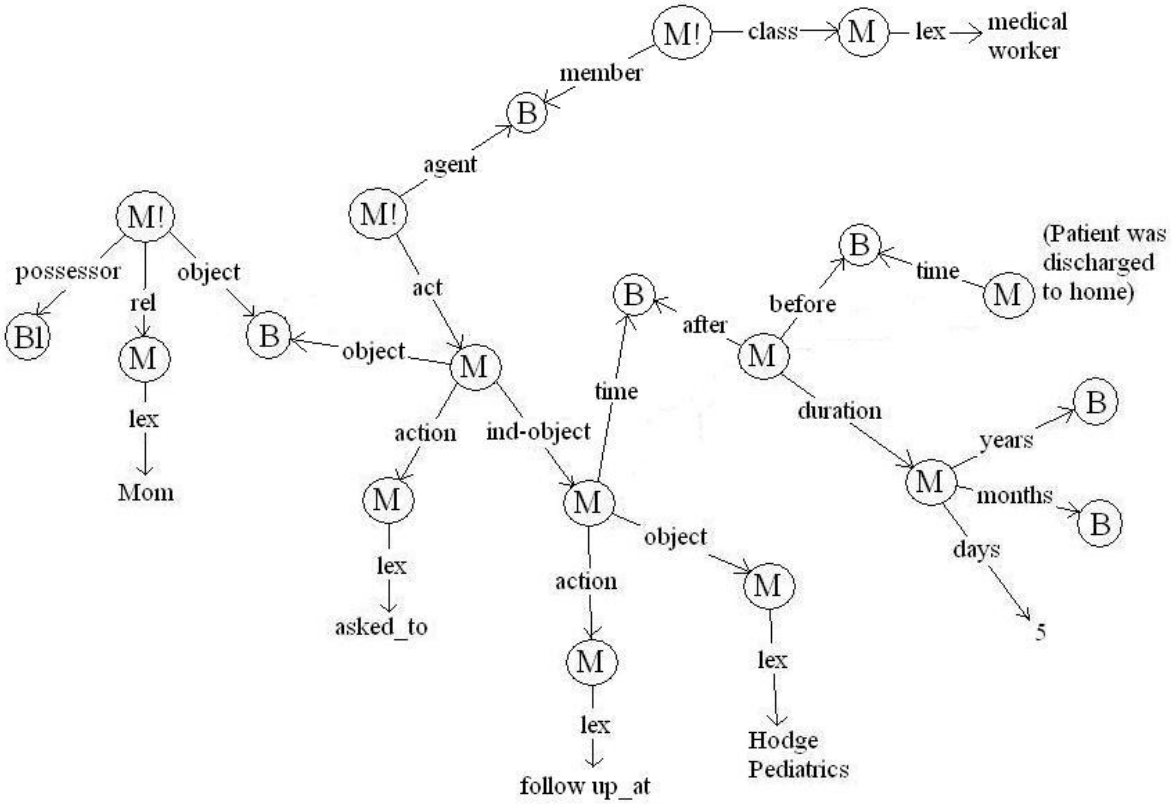
10. Patient was discharged to home with a diagnosis of RAD and was to continue albuterol aerosols and to complete a five day course of Prelone and Zithromax.



11. A home nebulizer machine was arranged for the patient.



12. Mom was asked to follow-up at Hodge pediatrics five days after discharge.



Appendix D: References

Stuart Shapiro, William Rapaport, et al (1996), “A Dictionary of SNePS Case Frames,” found at http://www.cse.buffalo.edu/sneps/Manuals/dictionary.pdf .

Shapiro, Stuart and the SNePS Implementation Group (1999), “SNePS 2.5 User Manual,” found at http://www.cse.buffalo.edu/~jsantore/snepsman/ .

Almeida, Michael (1995), “Time in Narratives,” Deixis In Narrative: A Cognitive Science Perspective, ed. Judith F. Duchan, Gail A. Bruder, and Lynne E. Hewitt.

Chun, Soon-Ae (1987), "SNePS Implementation of Possessive Phrases", *SNeRG Technical Note 19*.