Towards a natural spoken language order-entry system for the ICU:

Developing a language model from handwritten physician orders

By

Lara Fournier, B.S.

A Thesis

Presented to the Division of Medical Informatics and Outcomes Research

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of Master of Science Degree

September 2001

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Masters thesis of

Lara C. Fournier

Has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Member

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ACKNOWLEDGEMENTS

There are many people who have inspired, encouraged and supported me through this effort. The effort I describe is not only this thesis but also what it represents: the completion of this program and my Master of Science Degree in Medical Informatics. So I really owe an overall debt of gratitude to all the faculty and staff in the Division of Medical Informatics and Outcomes Research at Oregon Health & Science University, headed by the incomparable Dr. William Hersh.

First and foremost, I want to thank the members of my thesis committee. Dr. Judith Logan, the chair of the committee, used just the right balance of encouragement and forceful pushing to get things moving. Drs. Joan Ash and Zoë Stavri, whom I've had the pleasure of working with on several projects over the last two years, have been important motivators and mentors for me and provided valuable qualitative and linguistic expertise. Dr. Dale Kraemer, as usual, provided welcome statistical expertise but also the invaluable support as a friend by often allowing me to vent some of my stress by listening to that crazy singing fish he keeps in his office, and by just always being there. Dr. Deborah Eldredge was one of the first people with OHSU ICU experience to read through my proposal over a year ago, and it was her initial enthusiasm that helped bolster my resolve. I also thank her for coming through, on short notice, to become a valuable member of my committee.

I couldn't have achieved what I did without the generous help given by my principal OHSU ICU contact, Laura Criddle, CNS, who introduced the study to the clinicians and paved the way for me to proceed. She also provided critical clarification when I was struggling to interpret individual clinicians handwriting and abbreviations.

There are also several folks in the OHSU Medical Records Department without whose help I would not have had any data. They are Christine Slusarenko, Patty Todd, Stan Geier, and Tom Wright. Thanks so much!

Finally, the following people helped, provided expertise, encouraged or supported this work in some way: Mary Lavelle, Jim Carpenter, Paul Gorman, Mike Riben, Paul Ash, Pat DeGarmo, Dan Baker, Keith Klopfer, Susan Shapiro, Michael Grasso, Sharon Oviatt, Bill Smith, Jennifer Lai, Anne Cook, Mom, Dad, Jo and Jo, Dave, Shannon, Manish, Tim and Sharon. Thanks so much!

ABSTRACT

The work presented in this thesis explores one aspect of the medical documentation challenge: medical orders. The project is a first step towards examination of the use of speech recognition technology for order entry. A computer script, written in PERL, was created and tested using a random sample of handwritten orders from several Intensive Care Units (ICUs) at Oregon Health & Science University Hospital, with the goal of categorizing the orders into logical groups. The script searched for a distinct number of "keywords" that would allow categorizing of the orders by type, such as Medications or Imaging. The premise is that speech recognition accuracy improves when one limits the number and type of words the system is required to recognize. Handwritten orders were used because they were readily accessible and reflect a relatively unrestricted use of language, similar to a spoken interface. Randomly selected orders (1230) were obtained from 34 randomly selected patients. 605 of these were used as a development set. The remaining 625 served to test the keyword scheme developed. The results of this test revealed that the designed categorization scheme had an overall sensitivity of 83 % and a specificity of 93%. The study also revealed numeruous complexities to ICU orders.

INTRODUCTION

Background and Significance

Timely and accurate documentation of medical information remains a formidable challenge in health care. This challenge has persisted despite rapid advances in computer technologies. Technology infusion into medical practice has been gradual, the slowness largely due to the significant complexity of medical work. This complexity can be represented as an intricate web of collaboration, communication, and situational awareness. This web manages a vast and disparate amount of information and holds together with amazing strength when considering the high-paced and chaotic environment in which it exists.

Societal demands, increasing government regulation and media pressures are steadily adding to the burden of healthcare workers at all levels [1], thus creating undue pressure on the web of medical collaboration, communication and awareness. Ironically, introduced computer technologies intended to alleviate these burdens have, in many instances, worsened them [2]. While the belief remains that computer technologies need to play a role in medical practice, past implementation failures have shone a light on the vast complexity of medical work and spawned a realization that its integration with computer technologies is no simple task [3].

Placed in this context, speech recognition technology has potential benefits, that are enticing, especially when one imagines a future of talking to a computer as casually as to a colleague at work. Although research in this area dates back to the 1950s, only recently have advances in speech recognition systems seen success outside the laboratory [4, 5].

Speech Recognition

The distinguishing features of many speech recognition systems are speaker dependence, speech continuity, and vocabulary size. Speaker dependence reflects the system's ability to recognize an individual voice. A speaker-dependent system requires the user to enroll or train the system to his or her voice, by repeating phrases over a period of time until the system can build a template of that user's voice. Recently, speaker-independent systems have emerged that store voice templates from various speakers, thus enabling the system, within limits, to recognize speech from many speakers without prior enrollment or training [6].

Speech continuity reflects how the user speaks to the system and can be either discrete or continuous. Earlier systems were discrete, requiring users to "pause…between…words." More recently, technological advances have spawned continuous speech systems that allow users to speak more naturally without pausing [6].

The vocabulary of a speech system is the list of words the system is able to recognize. The larger this vocabulary, the more complex the computational-system requirements must be; the vocabulary, therefore, has an effect on system speed, recognition accuracy and hardware requirements [6].

Speech systems can be further classified by their functionality. One of the predominant functions used in the United States is that of command and control, where spoken words, the commands, are used as a substitute for pointing devices or buttons to control the computer functions. These systems generally operate with a limited vocabulary. The other function of speech input is to serve as a "voice typewriter," where bodies of text are created using voice input rather than a keyboard. This functionality generally requires a much larger vocabulary [7].

Speech systems are computationally intensive, generally operating on desktop workstation environments with a noise-filtering microphone, 128 MB RAM, and the fastest processors available (300 MHz or higher) [8]. Speech recognition has also made its way into the mobile phone arena, with various "voice-portals" used to browse the Internet. Similar applications are soon to be available for handheld devices as well [9].

Medical Applications

The early applications of speech recognition in medicine were disappointing. The systems were slow, in part because they were discrete, forcing the users to speak unnaturally (with pauses), and had a recognition rate of 30-40 words per minute at best. When adding in the time taken to correct frequent recognition errors, it was often faster to use the keyboard than speak. Not surprisingly, these early systems were not readily accepted [10]. In some radiology labs, users complained that the editing task distracted them from their job of image interpretation [11].

Some of these issues were resolved in the early 1990s with the advent of continuous speech systems that allowed users to speak naturally without pausing. Studies revealed that when these systems were combined with windows-based graphic control and feedback mechanisms, they were effective for certain medical applications. However, the effectiveness decreased when the grammars became very complex [12].

Developers of medical speech systems have targeted their efforts in domains where the vocabulary of the users follows predictable patterns, such as radiology reporting [13]. Restricting the vocabulary that a speech system must recognize helps insure a higher degree of accuracy [6]. These systems function in the role traditionally served by dictation and transcription. Medical applications of speech recognition technology outside the dictation/transcription function have not been widespread.

Speech recognition is likely to gain renewed attention in the near future with continued pressures from managed care and government regulation, each requiring that health care professionals document more patient encounters in less time. Adding to these challenges are mandates that institutions maintain clinical data in a format that facilitates research, analysis and quality management [1]. These conditions have caused many researchers and organizations to look for tools and technologies that can enable overburdened health care professionals to both comply with these mandates and improve productivity [14]. Some expect speech recognition to be one of those tools.

One of the most prominent benefits of speech recognition technology is an overall decrease in report turnaround time, compared to traditional transcription methods [11, 14]. It also allows for clinical findings to be documented closer to the time of detection [15, 16], and thus gathering data closer to the point of care. Decreasing the intervals of time between information generation and its subsequent documentation can support increased quality and accuracy of that information.

Speech input can also facilitate user navigation within a computer interface by providing an easy and quick way to access commands and options not readily seen on the screen [7]. This feature has clear advantages when considering the limited display size of increasingly popular handheld devices.

Speech recognition technology is also perceived as having the potential to overcome one of the obstacles to a fully computerized medical record: the direct capture of physician notes [10]. Some studies have indicated that health care providers would more readily accept computerized medical applications if speech were the method of input [17]. If true, this could improve acceptance of physician order entry (POE) systems, providing a link to decision support tools and the proven benefits of computerized reminders and alerts [18].

The excitement surrounding speech recognition technology has frequently led to misunderstandings and unrealistic expectations regarding the current capabilities of this technology. These misunderstandings have often led to abandonment and/or return of the speech-recognition product to its vendor. The emerging reality is that voice-mediated human-to-machine communication is not as simple as just replacing a keyboard or mouse with a voice command [19]. Successful human-computer interfaces using voice will likely need to be multi-modal, combining voice with screen feedback, pointing devices, gesturing, and handwriting [20].

Another possible explanation for why speech recognition has not been readily accepted into the clinical routine is because the speech systems used are often poorly integrated into the existing medical-documentation scenario [21]. When one large urban emergency department abandoned its voice-activated computer-based medical record, a retrospective analysis revealed many issues, only some of which were directly related to voice recognition. Too much time was required to dictate a case and proficiency training was lengthy. Broader issues included lack of adequate time for training and poor integration into existing work patterns, that is, when and how documentation occurs among other important tasks [22]. The needs of all users should be carefully examined.

Improvements in speech-recognition technology and language-processing techniques are still needed before this input modality becomes widely accepted for clinical applications [17]. Research so far has illuminated the need for further study in the area of spoken human-to-machine communication. Examples include evaluating users' problem-solving behavior, usability of speech with other input modalities, language used, and task performance [20]. Even when significant advancements in speech-recognition technology are realized, it may become evident that some constraint of users' language is still needed.

Modeling the language used in a particular domain can serve to provide some constraint on the vocabulary by identifying patterns, categories and themes. Constraining the vocabulary in this way limits the number of words a future speech system must recognize and thus could improve overall accuracy of the system by minimizing recognition errors.

The impetus for this thesis project is to evaluate an order entry language model for a spoken language interface in the absence of an operational speech recognition system. The expectation is that improvements in speech recognition technology are ongoing. In the interim, value will be gained from examining the broader issues of how, where, and when spoken language input could fit into clinical routine. This project attempts to examine these issues from the singular domain of an Intensive Care Unit.

Setting: The Intensive Care Unit

Intensive care units (ICUs) generally care for patients with a broad range of clinical conditions involving dysfunction of one or more organs, particularly respiratory and cardiovascular systems. These patients require intensive monitoring and often need some form of mechanical and/or pharmacological support [23].

Despite, or because of, abundant technologies, intensive care is expensive, labor-intensive, and people-oriented, requiring the management of patients, families, staff, consultants, technology and the expenditure of resources. The responsibility for this management generally belongs to the physician and nursing leadership of the ICU [24].

In addition to managing the patient, clinicians (physicians and nurses) manage the flow of information and monitor the performance of the hardware and the entire process of care. These roles are interwoven. It is unclear, however, whether they are separable, or if so, whether they should be [24].

Studies have supported the importance of collaboration between the physician and nurse as care providers in intensive care. Input provided from both professions can produce decisions based on more complete information, which, in turn, can lead to better patient outcomes [25].

Microprocessor advancements have brought sophisticated instrumentation and machinery to the bedside in modern ICUs. Increasing the sophistication of the care can also lead to increased complexity. As new therapies become accepted and widely available, the added complexity not only creates beneficial opportunities, but also increases the potential for error [24].

One of the formidable challenges for ICUs is how to integrate computerized decision support into clinical decision-making. This task requires converting a great deal of data into information that is malleable, accessible and useful to clinicians. While some efforts have been made in this direction, it still remains an elusive goal [24].

The ICU, therefore, represents a highly complex and novel setting in which to explore the potential of speech-recognition technologies. If carefully studied and integrated with the workflow, speech recognition could become a powerful tool to clinicians in ICUs to aid in the tasks of managing patients, information and communication.

Possibilities for Speech Recognition in the ICU

One of the obvious advantages to a spoken language interface is the ability to input data while hands are occupied with other tasks. Without speech recognition, the task of documentation in these situations has to come afterwards. The ability to input data at the point and time of care, using a spoken interface, could allow more timely documentation, maintain quality of the data, and provide instantaneous feedback if linked to decision support.

In the future, it is likely that handheld computers will be able to process spoken input. Such an advancement could spawn a multi-modal tool that enhances and streamlines existing collaborative processes and communication within the ICU.

Picture the following scenario:

An acute-care nurse practitioner converses with a physician during morning rounds about a patient who has adult respiratory distress syndrome (ARDS). They discuss changing the patient's ventilator settings. The physician asks the nurse some questions about the patient's status, including lab results. The nurse activates her handheld *computer to receive spoken input and says "Lab results…John Doe," which, by default, brings up the most recent lab results. As they discuss the lab results, the physician, anticipating that he will need to enter new orders for the patient, uses a stylus to quietly get into the order entry module on his own handheld device. Through discussion, they come to an agreement. The physician then activates his handheld to receive spoken input and repeats the order into it, still in the presence of the nurse. He gives her a look to see if she agrees with what he just said, and she nods. He reviews and electronically signs the order. Over a wireless network, the order is instantly documented and communicated to the appropriate staff.*

In this scenario, a tiny headset-microphone, worn by physicians and nurses, could allow for easy voice input into handheld device. Such microphones could, perhaps, become part of the standard healthcare attire similar to a stethoscope. The potential users of such a tool could include anyone in the ICU who interacts with information systems, whether that system is paper, electronic or human.

Since professional collaboration is a key feature of the ICU environment, such a voice-enabled tool as depicted in the previous scenario should enhance and support this collaboration rather than replace it. Computerized decision support integrated into an order entry system could provide feedback in the form of allergy checking and clinical practice guidelines, and thus become a contributor to decision making. It could either confirm what the clinician was thinking or bring critical information to the clinician's attention as necessary. The collaborative environment also requires a flexible tool that is multi-modal. The user should have the option to switch to different input modalities (as the physician did in the above scenario) when the situation warrants and the user sees fit – such as when someone else is talking or in the presence of a patient's family.

In a world where information systems are constantly changing and "improving," keeping up with the commands and functions of different computer systems presents a significant cognitive load on the user. In medicine, it is not uncommon for clinicians to need to learn how to navigate as many as five different information systems [26]. Keeping abreast of current medical treatments is already a formidable challenge for clinicians. A spoken language interface, a natural language interface in particular, has the advantage of broad familiarity that is virtually second nature – thus minimizing the cognitive load of memorizing commands. If a spoken input tool could interpret an utterance correctly most of the time, it also has the added benefit of saving the users' time by not forcing them to navigate and scroll through a myriad of screens and options.

Research is already underway to provide a natural-language interface to order entry processes. Researchers at the (U.S.) Department of Veterans Affairs Puget Sound Health Care System, Seattle Division, have implemented and evaluated an alternative ordering pathway known as "JIL" to replace the traditional menu-driven pathway in their Computerized Patient Record System (CPRS). Orders are typed in the user's natural language and processed using natural-language-processing techniques. The user's typed order is analyzed using partial-string pattern matching and proximity scoring, which then returns to the user a list of about 10-20 orders, sorted by their proximity scores. The user selects the intended order from this list and the request is then sent to the CPRS in the same manner as in the menu-driven system.

While JIL is not actively used with CPRS, an evaluation of the system was performed using 16 physician volunteers from the Puget Sound VA, Seattle Division. Results showed that all but three of the physicians were able to enter orders faster using the JIL system (mean time spared 16.06 ± 4.52 minutes; compared to CPRS's 17.69 ± 6.77 minutes; $P = 0.029$). Survey results also indicated that users found JIL easier to learn than the traditional, menu-driven CPRS order entry system. Interestingly, when compared with the task of handwriting orders, respondents indicated that the menu-driven system was much slower and more difficult to use, whereas they reported no significant difference with the natural-language interface. The researchers suggest that command-line (typed) entry of JIL could be replaced by voice recognition in the near future. Overall, the natural-language interface was well received by clinicians who found it more similar to their "usual" way of entering orders [27].

Rationale for this Study

The work at the Puget Sound VA has shown that a natural-language interface can be successful with clinician users. The logical next level to consider is a *spoken* natural-language interface. While it is unlikely that a natural-language speech-recognition device will be able to emulate a human's complete range of language understanding, such systems have been successful when accepting inputs about a limited range of topics [28]. When considering the focus of this study (ICU orders), it is the assumption of this researcher that if spoken orders could initially be categorized into various domains (topics) such as Medications or Imaging, accuracy of a potential speech-recognition application would improve while still maintaining a natural-language interface. This is because the system could then recognize the words contained in the spoken order using a more restricted vocabulary that pertains only to that category. If successful, this also holds potential for enabling speaker-independent speech recognition, where conceivably any speaker's voice could be recognized without prior system training.

Misrecognized input is a significant problem for spoken interfaces. Some researchers have addressed this issue by adjusting the model of the dialogue used. Using an automated patient history-taking system called Q-MED, researchers at Stanford University Medical Center modeled a dialogue where the computer asked the user (a patient) open-ended questions with large vocabularies or more directed questions with smaller ones. The system would switch to more directed questions if the user failed to answer the questions in the open-ended mode, or if the system misrecognized the input. This dialogue model allowed the user to use more natural language during the interviewing process. An evaluation of Q-Med, using 200 predefined sentences, yielded an overall semantic accuracy of 87 % [29].

Figure 1 provides a visual representation of a language model for a potential speech-recognition system that forms the premise for this thesis project. Levels 1, 2 and 3 represent different steps in the process of recognizing a spoken order, in this case, a medication order for Lovenox.

In Level 1 of Figure 1, a clinician speaks an order into a system (e.g., a handheld device or a desktop workstation with microphone): "Lovenox 40 mg subQ q daily." Acknowledging that medical orders, especially in an ICU, could represent a vast vocabulary of words, the *initial* goal in Level 1 is to recognize only the presence of a limited number of keywords and not all the words in the spoken utterance. These keywords, if present, would allow the spoken order to be placed into the appropriate category. In this example, the order type is "Medications," recognizable by the keyword "SubQ."

Further recognition of the spoken order can now take place, in Level 2, using a restricted vocabulary with words relevant solely to medication orders, and using knowledge of relevant grammar. For example, the "Medications" vocabulary in Level 2 could consist of all words used to represent dose (e.g., *mg, mEq, mmol…*), route (e.g., *IV, SQ, PO…*), frequency (e.g., *q 4 hrs, tid, bid…*) and duration (e.g., *x3 days, x2 weeks…*). Depending on the complexity of the vocabulary and grammar for a category, complete recognition of all other words, including drug names, may occur in Level 2.

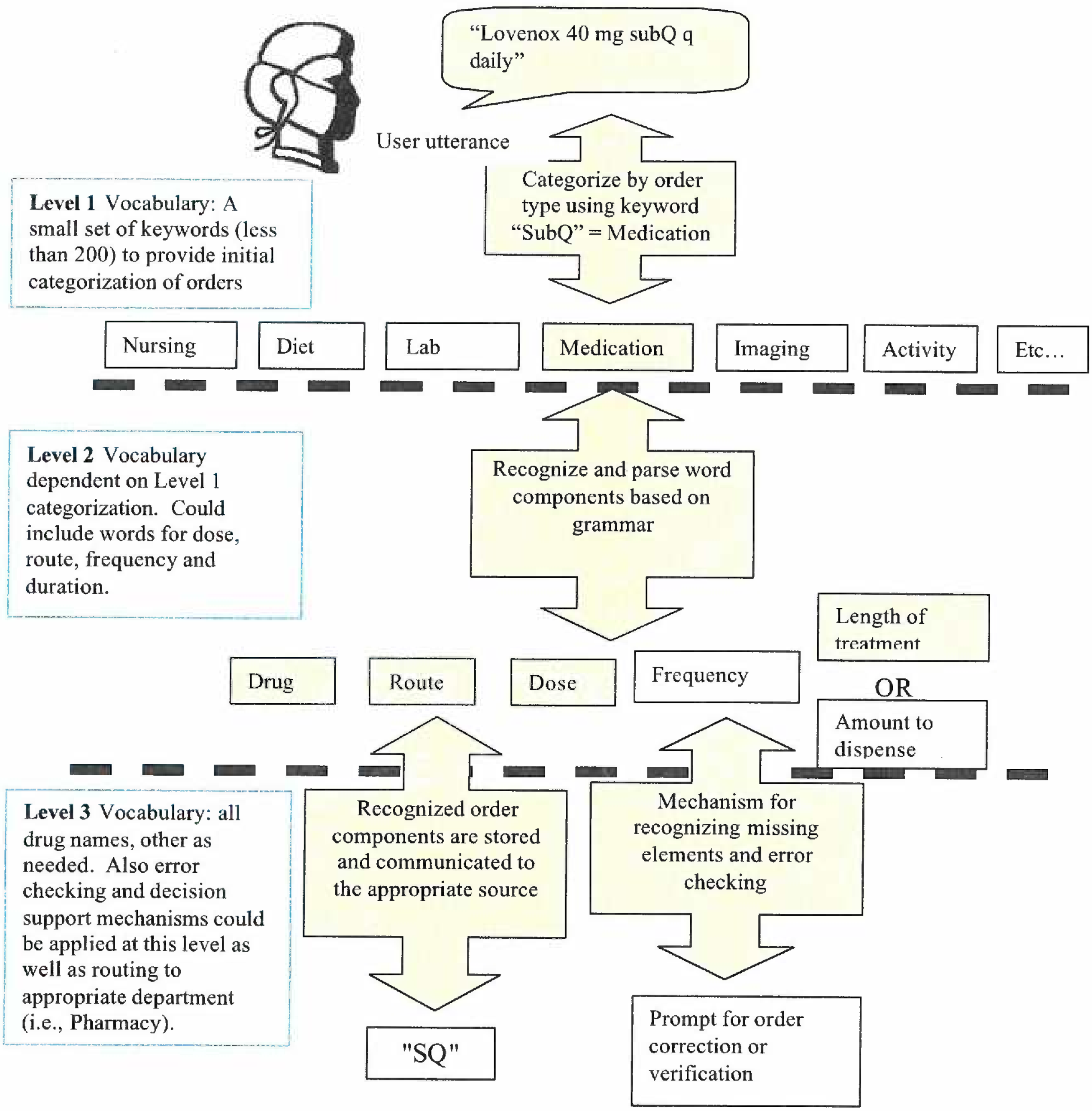


Figure 1. Design of an ICU order-entry language model. This study deals with Level 1 – mapping orders to initial categories.

Level 3 in the model allows for error checking and possible decision-support mechanisms to be fed back to the user. It also provides another level, if needed, for recognition when dealing with categories that have complex vocabularies.

The scope of this thesis deals *only* with Level 1 of the model represented in Figure 1. This entails the development and testing of a keyword-based categorization scheme for ICU orders.

The Questions

This research asks whether or not ICU orders can be categorized using a limited vocabulary of keywords. If so, then this project represents a first step in developing, for the ICU, a physician order entry system which could answer the larger question: Can a computable language model be determined for ICU orders? There are, however, two other questions that must be asked, questions which both determine the results and may be determined *by* the results of this project.

- *Is the ICU too complex an environment for development and use of a speech-recognition order-entry system?*

Developers of speech systems have generally steered away from intensive care units because of the complexity of the work and vocabularies there[30]. This study will give further validity to (or help invalidate) this conern; which is of particular significance because ICU nurses and physicians have expressed considerable interest in speech technologies [31].

- *Can a language model of ICU orders be created from handwritten physician orders?*

Handwritten orders were chosen because they reflect a relatively unconstrained use of language and are more abundant and readily attainable than audio recordings of verbal orders, and therefore represent a good starting point for study.

Specific Aims

The specific aims of this project are:

♦ Aim 1: *Develop a Level 1 model of the language of ICU orders.*

Consult with domain experts (ICU physicians and nurses) and analyze handwritten ICU orders (the Development Set); build a Level 1 keyword list to classify orders by type. Create a PERL script using the Development Set that parses the file, searches for the keywords, and categorizes orders.

♦ Aim 2: *Test the Level 1 model*

Test the validity of the model on a second set of handwritten ICU orders (the Test Set). Measure the percent of orders in the Test Set which can be correctly classified when compared to expert opinion [32].

♦ Aim 3: *Report the findings*

Analyze the results of the test to ascertain the validity of the model and attempt to answer the questions integral to the project. Target measures are the percent of orders correctly categorized in the Test Set, the percent correctly categorized within each category, and the percent of the orders in the Test Set that failed to go into any category.

METHODS

The overall methodology used in this project is generally described above and will be detailed below as three steps: data collection, development phase, and evaluation.

Step 1. Data Collection

Data collection was systematic. The target population was all orders written and interpreted in English from any ICU in any hospital. The accessible population is all Oregon Health & Science University (OHSU) ICU orders. The intended sample is all OHSU ICU orders from patients admitted during a three-month period (November 1, 2000 through January 31, 2001). The actual sample is a subset of those orders, specifically 1230 orders that comprise the development and test sets.

Before data collection could begin, the research proposal and required forms were submitted to the OHSU Institutional Review Board (IRB). Since the emphasis of the project is on the range of words (vocabulary) used in handwritten physician orders, and on developing a keyword list that represents and classifies ICU orders, there was no need for patient information. In the paperword submitted to the IRB it was explained that all patient-identifying information on order sheets would be blocked out before photocopying, and that the researcher would retain the photocopies only. In addition, this research did not attempt to evaluate the quality of orders. To insure that orders collected represented a variety of physicians, some basic demographic information on the physicians was collected, including job title and department, with the assistance of the

OHSU Medical Records Department. These procedures were clearly outlined in the IRB paperwork and the study was allowed to proceed under a "consent exempt" status.

The data collection consisted of three phases. Phase 1 involved deciding on the number and type of ICUs to be included in the study and then selecting a sample of patients admitted to those ICUs in the designated time period. Phase 2 involved making photocopies of the handwritten ICU orders found in the selected patients' medical records. Phase 3 involved taking a random sample of those photocopied orders to achieve the goal of approximately 1000 orders. The goal of 1000 orders was chosen because it seemed large enough to perform analyses but still feasible to handle (e.g., collect and transcribe into files) in the study time frame of an academic year. The researcher then transcribed the orders into electronic text files that could be processed through a parsing script written in the PERL scripting language [33]. Approximately half of those orders served as a development set to refine the language model. The remaining half was set aside to test the model once the development phase was finished.

Phase 1 -- Sampling

One of the goals of the research was to develop a keyword-based language model that was robust and capable of handling many different kinds of orders. To keep the scope of the project manageable, two ICUs deemed most likely to have a broad variety of orders, were selected. After consultation with a collaborating domain expert (an OHSU ICU clinical nurse specialist), it was determined that the Cardiac-Medical ICU (CMICU) and

Trauma-Neuro ICU (TNICU) were most likely to provide the needed variety. As noted below, however, this plan was slightly broadended.

A list of patients admitted to the CMICU or TNICU over the designated 3-month period (November 1, 2000 - January 31, 2001) was obtained from OHSU's Infection Control Department lists of *newly* admitted patients. The list obtained for this time period contained approximately 600 patient names and medical record numbers. 100 patients' names and medical record numbers were selected and given to the Medical Records Department so their records could be retrieved from storage. These were selected by taking every 5^{th} name on the list until 100 patients were selected.

Phase 2 – Obtaining ICU Orders

The researcher examined the records retrieved in order to determine where ICU orders began and ended. Often this was not a simple task. While the order sheets for hospitalization are clearly marked, there is no clear-cut convention followed on the order sheets to identify the movement of a patient into or out of an ICU.

In most cases, the start of ICU orders could be determined by a page of admission orders with the first line reading similar to the following: "Admit to 7A CMICU…" Photocopying began with this page and continued until a page with a transfer or discharge order was found. Many of the patients had more than one admission to an ICU, which necessitated that every order sheet be examined to make sure that all their relevant ICU orders were captured. In the instances where an admission order to an ICU could not be identified from these orders, the researcher went to the other areas of the medical record (e.g., Progress Notes or Flow Sheets), to estimate the dates the patient was in the ICU. The order sheets corresponding to those dates were then photocopied with all patient identifiers covered. In many cases, the selected patients had been admitted to other ICUs (e.g., Surgical ICU's) as well as the TNICU and the CMICU. Because of the difficulties differentiating specific ICUs in those cases, all their ICU orders were copied too. This departure from the original plan should have no significant effect on the value of the data obtained.

Not all of the 100 patient records, selected for retrieval, were reviewed in this study. First, records needing further scrutiny by Medical Records staff were not readily available. Second, after six weeks it was determined that the 900+ handwritten pages of orders, from the first 56 patient records retrieved, was more than enough for a sample of 1000 orders, so no more records were reviewed. There was a wide variety in the number of pages of orders for each patient. Although some had only one page, a few had over 100 pages of orders.

Phase 3 – Order Transcription

Each of the 56 records' photocopied order sheets was given a letter code unique to the particular patient. After a rudimentary examination of several of the photocopies, an estimate was made of the average number of orders per page. It was estimated that a random sample of 5 pages out of a random sample of 34 of the 56 patient records should yield around 1000 orders. As detailed in Appendix A, 1230 orders were actually obtained with this method. (Five pages on 34 patients were chosen, as opposed to obtaining fewer pages on more patients, in order to get a sample that reflects more depth and continuity on an individual patient's care.)

After assigning a number (1-56) to each patients' set of orders, 34 order sets were chosen using a random numbers table. A second round of randomization was applied when selecting which five photocopied pages to use from the 34 order sets. Again, using a random numbers table, five pages were randomly selected from each of the 34 sets of orders. If the patient had five or fewer pages total, then all pages were used.

All selected orders were transcribed into an electronic text file, yielding a total of 1230 orders. Each order was then assigned a unique number. A duplicate of this file was made and in the duplicate, many of the abbreviations were spelled out and jargon explained. The original file was used for development and testing of the language model. References about verbal/telephone orders and chart checks were also left out of the original file. The file was then divided, approximately in half, to give Development and Test Sets. A total of 605 orders from 16 patients were in the Development Set, 625 orders on 18 patients in the Test Set (Table 1).

Data Attribute	Development Set	Test Set
Number of patients	16	18
Total Number of Orders	605	625

Table 1. Attributes of the Development and Test Sets.

Issues Encountered

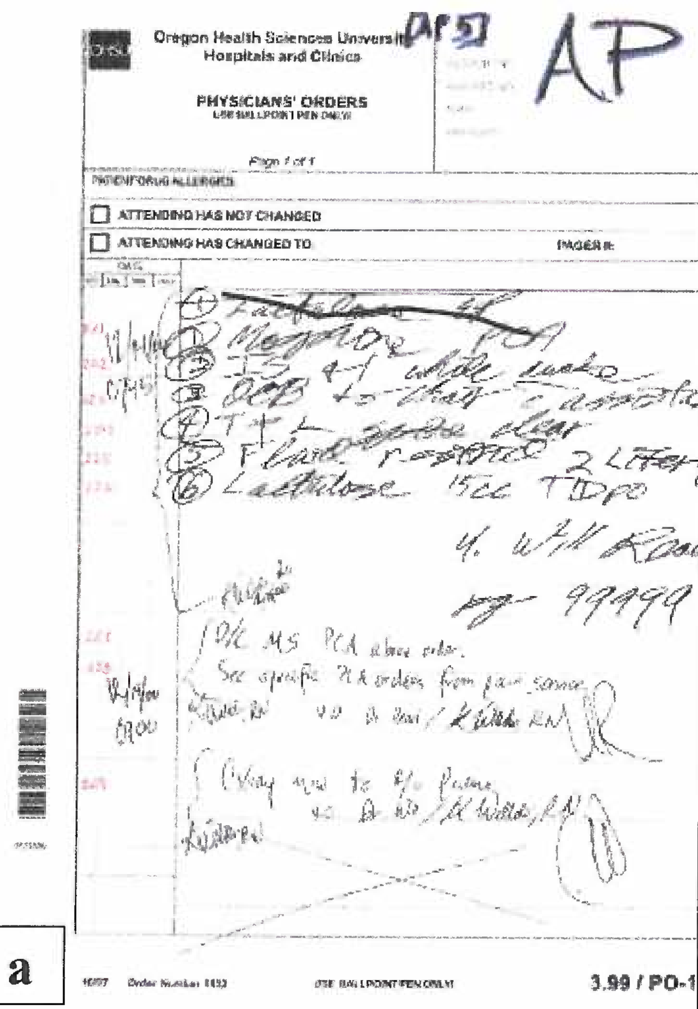
Soon after the task of typing the orders began, it became apparent that there would be instances where modification of the orders would be necessary, a process which can be termed "normalization", although the original text was used as much as possible. This was required in order to make the PERL script (used to develop and test the keyword-based categorization) easier to write and maintain. Numerous conventions were established and followed, are outlined in Appendix B. An example of such normalization is the convention chosen for the abbreviations of "milligrams" and "magnesium." In many instances, the handwritten abbreviations for these two terms involved the same two letters "mg" – sometimes the "m" and/or the "g" would be upper case, sometimes not. Based on the context of the order, the researcher had to determine which term an "mg" in a handwritten order stood for (milligrams or magnesium), and then follow the same convention for that term throughout the typing. This is justified because the ultimate intent of this project is to classify verbal orders, where these two words should be easily distinguishable. For this research, the term "milligrams" was always typed in lower case as "mg" and "magnesium" as "Mg" increasing the computability of the file. In most cases, the context was relatively simple to identify by the researcher. As in the above example, "milligrams" usually accompanies a number and a medication name. Magnesium, a laboratory blood test, was often written in a string with other familiar blood tests. In situations where the context was not clear to the researcher, help was obtained from domain experts.

In addition to the difficulties of deciphering individual handwriting styles, it should be noted that handwritten orders encountered in this research were rife with abbreviations. This forced the researcher to consult an abbreviation codebook or, failing that, with domain experts [34]. The researcher maintained a glossary of frequently encountered abbreviations, included here as Appendix C.

Another complex issue encountered early in this research was the determination of what constitutes a single order. For example, a line of text reading "CBC, Chem 7, Mg, Phos, this am and q am x 3 days" could be viewed as one order, as four separate lab orders (CBC, Chem-7 panel, Magnesium and Phosphate), or even as 16 orders when considering their repetition over four days. While there may be arguments for and against any chosen convention, this study required that some convention be adopted and used consistently.

In speech, an order might be an utterance - a sentence or a string of spoken words - where the start and end is marked by a brief silence or may be inferred based on more complex clues, such as the level of voice intonation [35]. It was assumed here that if the clinician wrote a series of orders in a natural grouping, then he would also be likely to verbalize those orders together in a single utterance. Generally, clinicians separate orders by writing them on a new line on the order sheet and/or by numbering them. So, the example "CBC, Chem 7, Mg, Phos, this am and q am times 3 days", written together on a single line, would be considered one single order.

Figure 2a presents a mock-up of what the raw data looked like and demonstrates some of the other problems encountered in this study, namely, poor handwriting and use of abbreviations. Figure 2b is a representation of the electronic text version of this order sheet. The "AP" represents the letter code assigned to a record by the researcher and all pages of orders copied from that record were numbered sequentially, ("[AP 5]" at the top center of Figure 2).



a

221. Morphine PCA [AP5]
222. IS x 1 while awake [AP5]
223. OOB to chair with assistance [AP5]
224. T+L spine clear [AP5]
225. Fluid restricted 2 Liter q d [AP5]
226. Lactulose 15 cc TID PO [AP5]
227. D/C MS PCA above order [AP5]
228. See specific PCA order from pain service [AP5]
229. CXray now to r/o Pneumo [AP5]

b

Figure 2 (a.) Mock-up of handwritten orders illustrating what the original data looked like, including some of the handwriting and abbreviation issues encountered (b.) A representation of the transcription of these orders.

Each order was also assigned a unique number, which was recorded next to the order in the typed files as well on the raw data sheet (the red numbers on the left side in Figure 2a).

At the end of each order in the electronic files was the letter-code pertaining to the patient, typed in brackets along with the page number that the order came from within that patient's order sheets. For example, the typed version of the first order in Figure 2 would be as follows " 221. Morphine PCA [AP5]". The number "221" represents the unique I.D. number assigned that order. This was done to aid in the retrieval process if the raw data needed to be re-examined.

Several conventions in order writing, incidental to this study, were observed. First, errors are generally crossed out with a single line and initialed (e.g., the first order in Figure 2). Second, at times, orders are written by a clinician but need co-signature by a responsible physician, such as orders given verbally to a nurse (bottom of Figure 2a, where "V.O." indicates a verbal order). These are written down by a nurse but later signed by the physician that gave the order. Also, in a teaching hospital, it is common that orders written by resident physicians require co-signature by the attending physician. Finally, as orders are carried out, a nurse reviews the orders and makes a sweeping mark, usually with a red pen, (does not show up on photocopy) around the corresponding orders. The nurse then puts his/her signature at the end of this mark and a note in left column as to the date and time the order was noted (see the left side of Figure 2a).

The Clinicians Writing the Orders

Some basic information regarding the clinicians who wrote the orders was obtained with the expert assistance of staff members from OHSU's Medical Records departments. They examined the signatures corresponding to the 625 orders in the test set and helped the researcher ascertain how many clinicians were responsible in writing the orders as well as their job titles and departments.

Step 2. Development Phase

From the selected sample of 1230 orders, approximately half (605) were used as a development set. This approximation was determined as the point closest to the true center (615) of the total orders without splitting up individual patient records. The researcher studied this development set, looking for keywords and categories to create a model that could then be represented in a program and tested. The development set was then processed using the program, in an iterative fashion, and changes were made to the keywords, categories, and program based on those results. This iterative process continued until as many of the orders as possible, were appropriately categorized.

Category and Keyword Selection

Selection of order categories was determined by studying the orders in the development set and through consultations with an ICU domain expert (OHSU ICU Clinical Nurse Specialist). Once logical categories were determined, the next step was to look at the words used in those order types and try to determine potential keywords that might uniquely categorize those orders.

The keywords in this language model needed to be relatively unique to a particular category. Keywords were selected by careful examination of the orders belonging in certain categories. To verify their success as keywords, they were embedded in the PERL script for testing. In many cases, keywords that initially appeared to be good candidates were eliminated when tested in the program. For example: The keyword "cc" (cubic centimeters) seemed to be a good candidate to identify medication orders, but after testing it in the PERL script, it was discovered that it is also used in Diet orders, so "cc" was eliminated as a possible keyword. Selections of all keywords followed similar examinations and are listed in Appendix D.

The PERL Script

The program used was written in PERL and designed to examine the text file containing the orders and, for each order, parse the individual words delineated by whitespaces. It would then look for matches for each word in the set of keywords. Once a match was found, the program would store the whole order into an array representing the chosen category. Orders were not analyzed after the first match was found. Each order was analyzed sequentially until the end of the file was reached. The program was case sensitive to maintain the differing semantics frequently represented by abbreviations. For example, "IS" in all-caps is an abbreviation that stands for "Incentive Spirometer," in lower case it's simply the verb "is." In addition to carefully selecting keywords, it became clear from initial test-runs of the PERL script, that the sequence in which the orders are analyzed in the program is important. To insure the best categorization possible, the sequence needed to be adjusted and fine-tuned. For example the lab order "CBC, BMS, Phos, Mg…when PICC cleared" was wrongly placed in the Nursing category because "PICC" was a keyword for Nursing orders and the program was searching for those keywords before it looked for Lab order keywords. To correct this problem and still maintain "PICC" as a Nursing keyword, the sequence in which the keywords were matched was reversed so that the keywords for Lab orders were searched for prior to the keywords for Nursing orders. Many similar adjustments were made to the program's analysis sequence.

Category and Keyword Breakdown

Sixteen initial order categories were identified as follows:

- Discontinued orders
- Medications
- Fluids
- Transfusions
- Laboratory tests
- Imaging
- Nursing assessments
- Activity
- Equipment
- Respiratory care

- Diet,
- Admit/Transfer/Discharge
- Consults
- Diagnosis
- Condition
- Allergies

After exhaustive work using the Development Set, it became clear that many of the orders were not categorizing properly into one of the 16 groups listed above, even after adjustments were made to the keywords and the sequence that the program looked for keywords. It became apparent that merging certain categories could solve the problem. Specifically, there was so much keyword overlap between Medication orders, Transfusion orders and Fluid orders that separating them was nearly impossible. Accordingly, the three categories were joined into one "Medications/Fluids/Transfusions" category.

Step 3. Evaluation Phase

The developed model and associated PERL script were then used to categorize the 625 orders comprising the Test Set. An expert panel of four clinicians (3 physicians and an OHSU ICU clinical nurse specialist) judged the results to determine which orders were correctly categorized. For the few orders where there was not a consensus, two experts (one of the original 3 physicians and a different ICU nurse) met and made a final determination of correct categorization.

The target measures were the percent of the total 625 orders that were correctly categorized, the percent of orders correctly categorized within each category, and the percent of orders that failed to go into any category. If the model is robust, the last measure should be minimal, although it was expected that some orders would not fit into the chosen categories, based on the Development Set work. The results of the categorization from the Test Set form the results of the study and help determine whether or not the model is an adequate representation of ICU orders.

RESULTS

Clinician and Order Attributes

Table 2 shows details regarding the clinicians who wrote the 625 orders in the test set. For the 18 patients whose orders were used in the Test Set, a total of 74 clinicians were involved, representing 10 departments or specialties. Physicians from General Surgery and Medicine were responsible for most of the orders. For the purposes of this research, clinicians were determined to be responsible when orders were written in their own handwriting or if they signed a verbal or telephone order. Orders written by residents, respiratory therapists or physician assistants may require co-signatures by supervising physicians. Since the focus of this study is on the words (vocabulary) chosen by the clinicians, the co-signing attending physicians were not considered responsible for the order.

Type of Clinician	Number of clinicians responsible for orders in Test Set
Resident	50
Attending	13
Nurse Practitioners	4
Respiratory Therapists	3
Registered Nurses*	3
Physician Assistant	1
TOTAL	74

Table 2. Types of clinicians responsible for the orders written in the test set. The clinicians came from 10 different departments/specialties with the majority representing General Surgery and Medicine. **Orders written by registered nurses were post-mortem orders for organ harvest.*

Table 3 shows the percentage of orders in the test set that were either verbal or telephone orders. OHSU policy dictates that orders of this type be clearly marked (V.O. or T.O.), so they were fairly easy to identify and count.

Verbal Order Type	Number of Orders	Percent
Verbal Orders	69	11.0
Telephone Orders	44	7.0
TOTAL	113	18.0

Table 3. Percent of verbal and telephone orders occurring in the test set.

Order Categories

The development process resulted in 9 category types (Table 4). In addition to merging the problematic categories, the Diagnosis, Condition, and Allergies categories were merged as well. The latter three were not problematic categories, but had relatively few keywords and were considered informational and not true orders, allowing these three to be merged for additional efficiency. Using this methodology, 94% of all orders in the development set could be successfully categorized.

Category Type	Definition of Category	Examples of Orders	Examples of Keywords
Discontinued Orders (D/C)	A stop to previous order. Can include orders from any category	D/C IV Zantac	D/C
Medications / Fluids / Transfusions	All drug, IV fluids, blood and transfusions orders.	1. Fentanyl 25 mcg IV x 1 2. NS (Normal Saline) at 70 cc/ hr 3. Transfuse 2 units PRBCs (packed red blood cells).	IV, mcg NS units, PRBCs
Imaging	Includes radiographic images, x-rays, EKG, ECG and ultrasound	EKG now and in am	EKG
Laboratory tests	All types of body fluid tests (mostly serum and blood) as well as microbiological tests.	CBC q 6 hrs x 24 hrs then q 8 hrs	CBC
Nursing assessments / Activity / Equipment / Respiratory Care	Broad category that includes many nursing related tasks including vitals, drain and catheter instructions, patient activity, equipment and respiratory care orders.	1. Nursing: strict IO's, daily weights 2. Call HO if T >38.5<35.5… 3. Bedrest 6 hrs post sheath pull 4. SCDs (sequential compression devices) at all times 5. O2 to keep Sat >= 95%	Nursing, IO's T Bedrest SCDs O2, Sat
Diet	Includes orders relating to food, nutrition intake and maintenance as well as feeding tubes.	Start Probalance at 10 cc / hr	Probalance
Admit / Transfer / Discharge	Orders outlining the movement of a patient, into or out of the ICU	1. Admit to MICU 2. Transfer 6CVA Blue 3. Discharge to Crestview SNF	Admit Transfer Discharge
Consults	Requests for evaluation or consultation. Can include: respiratory therapy, social services, nutrition and various specialists.	P.T. (Physical Therapy) eval / treat	P.T.
Diagnosis / Condition / Allergies	Informational non-orders that serve to guide the order writing process. Always a component of admission orders sets.	1. Dx: GI Bleed 2. Condition: guarded 3. Allergies: Sulfa	Dx: Condition: Allergies:

Table 4. Types of order categories defined with examples and samples of orders and keywords.

Appendix D has a list of all the keywords that the program parsed and used for categorization. While the program searched for variations of the same word, such as "Allergies:" or "allergy," this study classifies these as the same keyword. Only variations that might result in significantly different pronunciations were counted as separate keywords. Total resulting keywords for Level 1 of the language model (Figure 1) was 185.

Evaluation Results

Table 5 shows the distribution and percent of orders as categorized by expert judgment, which was considered the "gold standard" for this evaluation. The largest number of orders, by a substantial margin, fell into Medications/Fluids/Transfusions category (244). The next largest category was the Nursing assessments/Activity/Equipment/Respiratory care category with 122 orders.

Order Categories	Number of Orders	Percent
Discontinued orders	28	4.5
Medications/Fluids/Transfusions	244	39.0
Imaging	30	4.8
Laboratory tests	66	10.6
Nursing assessments/Activity /Equipment/Respiratory care	122	19.5
Diet	30	4.8
Admit/Transfer/Discharge	25	4.0
Consults	13	2.1
Diagnosis/Condition/Allergies	45	7.2
Uncategorized		
**Combination orders	20	3.2
Other	2	0.3
TOTAL	625	100.0

Table 5. Number and distribution of orders into categories based on expert judgment (N=625). ***The model did not include a mechanism for identifying and separating "Combination orders" which contained segments belonging to more than one category.*

Table 6 shows the results of the evaluation phase with the number of orders successfully categorized. Sensitivity of the model (Figure 3) is defined as the number of orders correctly categorized divided by the total number of orders that exist in that category as determined by expert judgment. Specificity of the model (Figure 4) is defined as the number of orders correctly categorized over the total number of orders the program placed into that category (both correctly and incorrectly). The PERL script did not have a mechanism for separating multiple orders of different categories ("Combination orders") so the expert reviewers were instructed to classify those as "uncategorized."

Order Category	# Orders by Experts	# Orders by Script		Sensitivity	Specificity
		# Correctly classified	*# Incorrectly classified*		
Discontinued orders	28	26	11	0.93	0.70
Medications / Fluids / Transfusions	244	231	15	0.95	0.94
Imaging	30	28	5	0.93	0.85
Laboratory tests	66	60	1	0.91	0.98
Nursing assessments / Activity / Equipment / Respiratory care	122	72	2	0.59	0.97
Diet	30	27	2	0.9	0.93
Admit / Transfer / Discharge	25	21	2	0.84	0.918
Consults	13	7	4	0.54	0.64
Diagnosis / Condition / Allergies	45	36	0	0.8	1
SUBTOTAL: Categorized Orders	603	508	37	.83	.93
Uncategorized					
"Combination" orders	20	0	0		
Other	2	1	74		
TOTALS	**625**	**509**	**116**		

Table 6. Results of the evaluation phase with sensitivity and specificity of model's keyword-based categorization of test set orders, by order category.

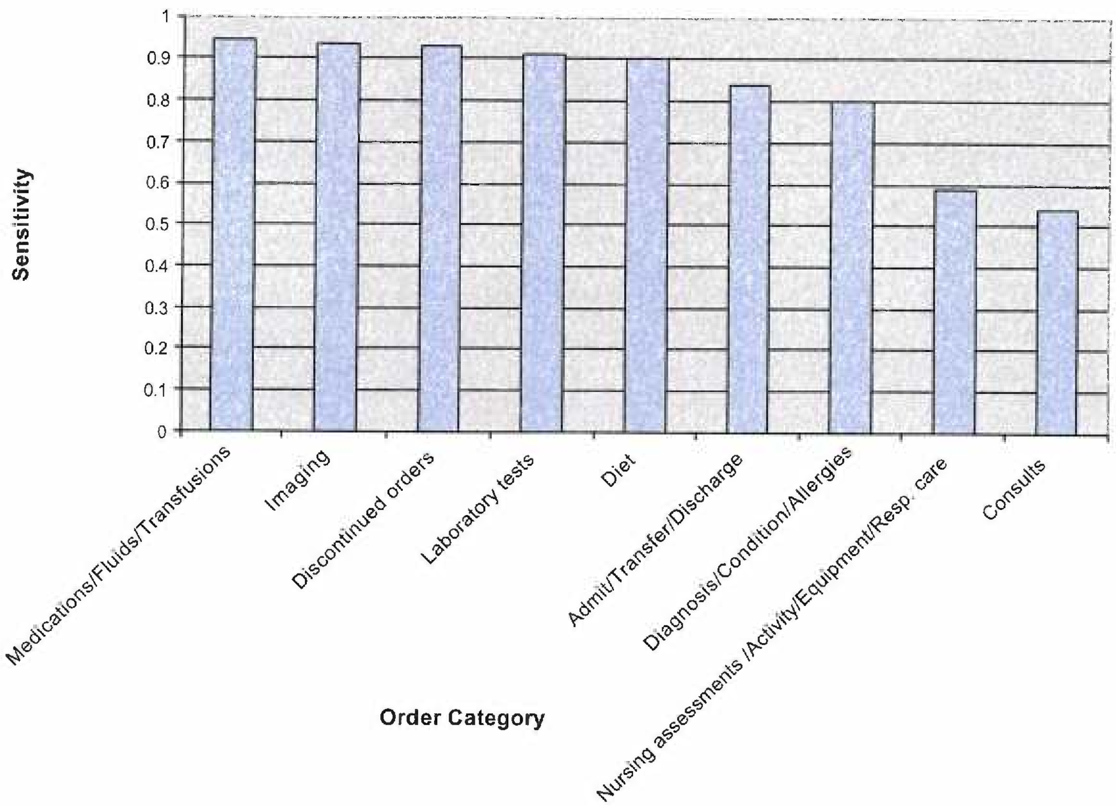


Figure 3. Sensitivity of the model based on the evaluation phase. Overall sensitivity was 83%. Sensitivity is defined as number of correctly categorized over total orders in that category (by expert consensus).

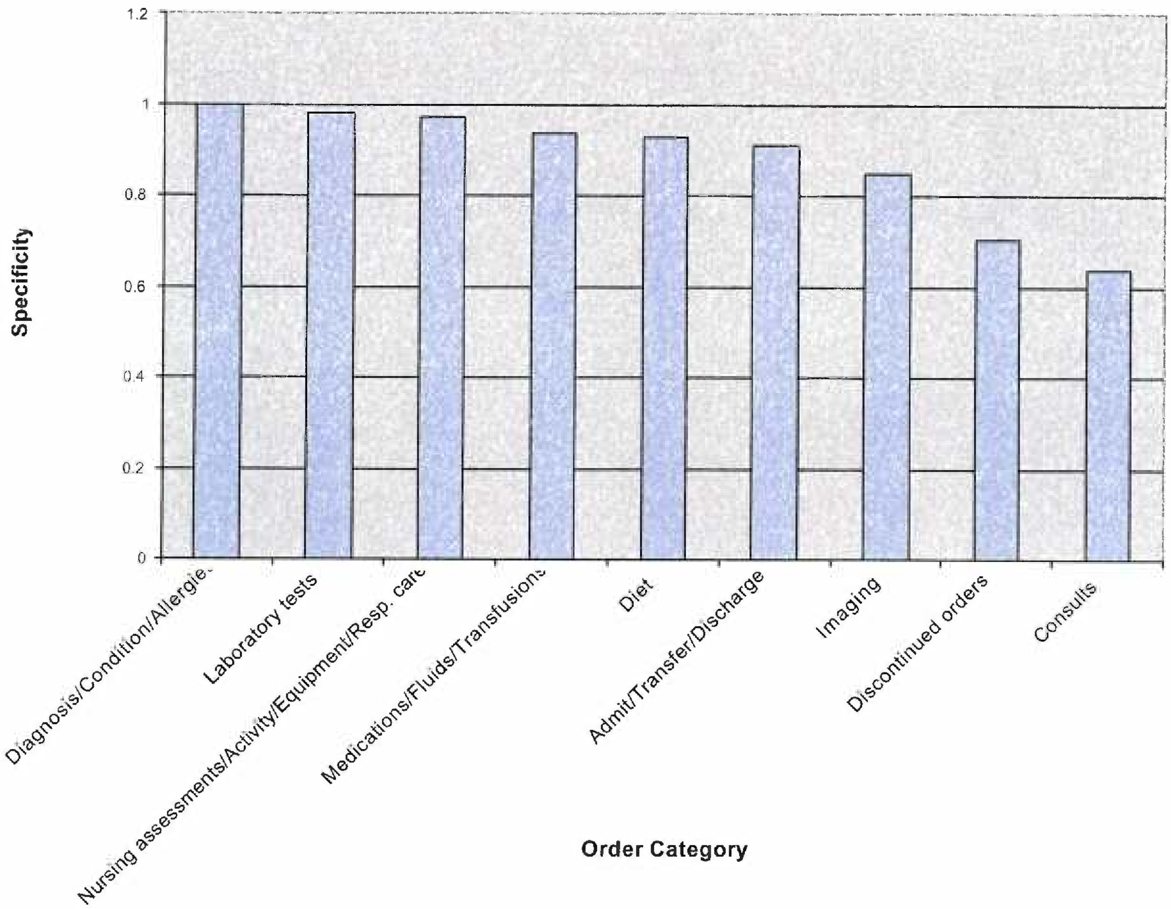


Figure 4. Specificity of the PERL script based on the evaluation phase. Overall specificity was 93%. Specificity is defined as the number of orders correctly categorized over the total orders retrieved for that category (correct or not).

The sensitivity was highest (over 0.9) for the Medications/Fluids/Transfusions, Discontinued (D/C), Imaging, Diet and Laboratory tests order categories, and lowest (below 0.6) for the Nursing assessments /Activity / Equipment / Respiratory care and Consults categories. For most categories, specificity was than sensitivity. The highest specificity (over 0.95) came from the Laboratory Tests, Diagnosis /Condition /Allergies and the Nursing assessments / Activity / Equipment / Respiratory care categories. The Admit / Transfer / Discharge, Diet and Imaging categories also had high results (over 0.9). Lowest specificity (below 0.8) results came from the Discontinued (D/C) and Consults categories.

A major problem not recognized until these results were obtained was the large number of "combination orders." These are defined as orders, which appear to be a single order (based on study's assumption of what constitutes a "single" order), but actually contain multiple segments belonging to different categories. The model and script were designed to look at vocabulary alone and did not have a mechanism for determining when segments of two or more different categories occurred in a single order. Also, each order could be placed in only one category. In addition, expert judgment was that only 2 orders from the Test Set (other than the combination orders) could not be categorized, while the model left 74 orders uncategorized (12%).

DISCUSSION

Although there were several issues relating to the complexity of the handwritten orders in this study that were not readily addressed in this keyword-based categorization scheme, it remains to be seen whether or not they will still be issues in a spoken language interface. There were 116 orders in the test set that were incorrectly categorized by this model and script, comprising 18.6% of the total data set of 625 orders. A number of these were "Combination orders" which resulted from an initial decision on the definition of an order (see METHODS above). This work also illustrates the complexity of many orders (i.e. a "Do this...then..."structure). Some of the orders could not be categorized, illustrating a problem either with the categories chosen or with the keyword selection for those categories.

All 116 orders, which were either incorrectly or not categorized, were closely scrutinized to determine why the error occurred. From this examination, six reasons were found to be responsible for incorrect categorization and they are shown, with examples, in Table 7 and Figure 5. The first and most significant reason for incorrectly categorized orders was that the order lacked at least one or more of the designated keywords, which comprised 66% of the total miscategorized orders. The next significant reason (comprising 17% of the total) was the previously mentioned combination orders, where two or more components of different categories occur in the same single order. Another 8% of the miscategorized orders were due to improperly selected keywords. This reflects the lack of an expert consensus process during the model development phase such that the keywords selected represented the wrong category.

In some cases, a single order of one category type contained extra keywords from other categories, included for informational purposes. In these cases, the extra keywords did not represent separate orders but still caused the order to be placed in the wrong category. These comprised 3% of the total miscategorized orders. Another 3% of miscategorized orders were due to the lack of the present model and PERL script to ascertain any context from the orders. This became a problem with abbreviations that can have more than one interpretation. The final and least significant reason for miscategorization (2 %) was the occurrence of a few orders that could not be placed into any of the nine designated order-type categories.

Examination of Problematic Categories

Using this framework, two of the categories with low sensitivities (< 0.6) were examined to gain insight as to why their sensitivities were so low. These were the Nursing assessments / Activity / Equipment / Respiratory care (abbreviated as "Nursing") and Consults categories. The sensitivities were .59 and .54, respectively.

The totals in Table 8 for the Nursing and Consult columns include orders of that type that were incorrectly placed into other categories, as well as orders of other categories that were placed incorrectly into the Nursing or Consult categories. This includes those that were part of combination orders. This differs, somewhat, from how the data is reflected in the overall distribution of orders in the first column in Table 6, where combination orders are kept distinct and not counted in the totals of the other categories.

Categorization Error	Description of Categorization Error	N* (%)	Examples	Causes
No Keyword	Order lacks one or more of the designated keywords	77 (66%)	• "Atrovent nebulizer per R.T." • "AP + Lat of (L) femur" • "Advance Dobhoff 4 cm"	• Keyword-set did not contain "nebulizer" because the word was not encountered in the development set. • Partial order-lacks the keyword "X-ray" • No Keyword exists due to improper order-type definitions; e.g. "Dobhoff" was not selected to represent Nursing Orders.
Improper Keyword Selection	Selected keywords define order type incorrectly	9 (8%)	"Finish current IV bag and hep lock IV"	Keyword set contained "IV" and placed order in Meds/Fluids/Transfusions category. However, expert consensus determined this order belongs in the Nursing Category
Multiple Keywords	Single order type contains keywords from more than one order category	4 (3%)	"Cx cath tip when D/C 'd"	While this Lab order contains the proper keyword "Cx", it was still miscategorized because it also contained the keyword "D/C".
Combination Orders	Two or more orders from different categories in a single order-utterance	20 (17%)	"CBC , BMS , ABG, CXR in am"	Single order was defined as what was written together on a single line. The PERL script had no mechanism for identifying and splitting up this string of four orders. Three belong to the Lab category while one, "CXR", belongs to the Imaging category.
Undefinable Orders	Orders that didn't fall into any of the 9 defined order types based on expert consensus	2 (2%)	"Vest - therapy's q 1 hr"	Taken out of context from the patient's entire set of orders, these particular orders were un-classifiable by expert consensus.
Context	Occurs when common abbreviations can have more than one interpretation.	4 (3%)	"D/C planning for midday"	PERL script had no mechanism for determining the context in which the abbreviation is used, e.g. to determine that "DC" means "Discharge" instead of "Discontinue" in this order.

Table 7. Six reasons for incorrect categorization by the model and script, along with sample orders and explanations for the error. **N = number of occurrences in the Test Set.*

However, when examining the problematic categories, the researcher felt that this additional information was important to the overall understanding of categorization error.

Type of Error	Overall		Nursing orders		Consult orders	
	N	%	N	%	N	%
No Keywords	77	66	39	65	9	75
Improper Keyword Selection	9	8	8	13	0	0
Multiple Keywords	4	3	2	3	0	0
Combination orders	20	17	9	15	2	17
Undefinable	2	1	2	3	0	0
Context	4	3	0	0	1	8
Total	116	100	60	100	12	100

Table 8. Number of Nursing and Consult orders incorrectly categorized by type of error and compared to overall error rates. **N = number of occurrences in the Test Set.*

Figures 5, 6 and 7 provide a visual representation of the information displayed in Table 8. Three pie charts represent the overall reasons for categorization error, Nursing categorization error and Consult categorization error. The similarites are clear. The the most significant causes of categorization error were due a lack of representative keywords, the next largest due to combination orders. These two problematic categories shared the same largest reason (No Keyword) for incorrect categorization due to inadequate development of the keyword list. The Nursing category, in particular, was lacking in its overall definition of the category and subsequently, in what kinds of keywords could represent this category. The results from the Consults category should be interpreted with some caution, as there were relatively few orders in this category to begin with.

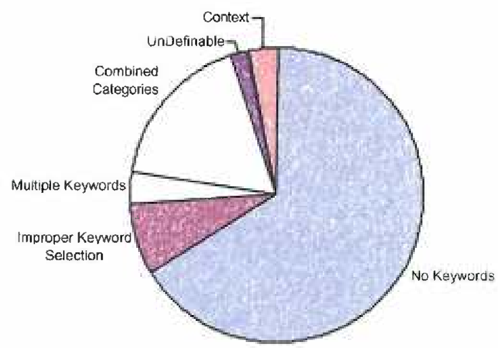


Figure 5. Reasons for overall categorization error

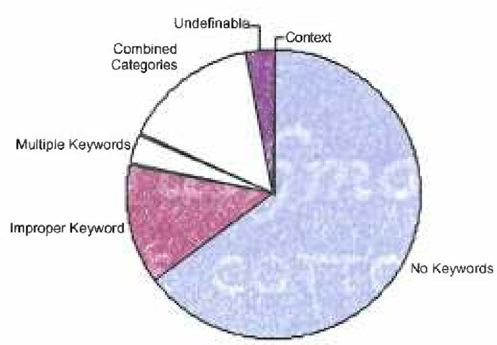


Figure 6. Reasons for Nursing categorization error

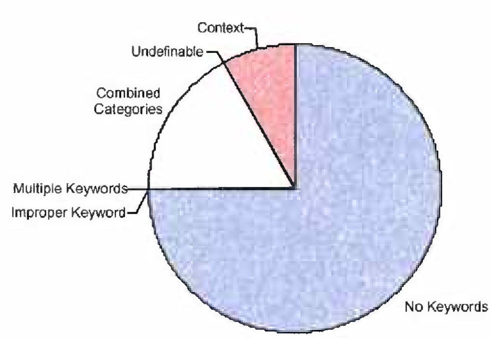


Figure 7. Reasons for Consult categorization error

Complex Orders

In addition to the reasons for incorrect categorization outlined above, other complex orders were observed. In this study, these have been identified as referent orders, conditional orders and modification orders. Orders of these types were not, by themselves, contributors to incorrect categorization.

Referent orders

Referent orders are defined as orders that make reference to other orders, often without much further details. These references could be to order sets (pre-printed or otherwise), to standard protocols or parameters, or to orders previously written for an individual patient. Referent orders may take advantage of the knowledge and expertise of the staff to whom the orders are addressed.

Examples of referent orders are:

- "Aspiration precautions as posted."
- "Post-op orders." and "Vitals: routine."
- "Please lighten sedation and run weaning parameters."

Some referent orders were incorrectly categorized because they lacked the necessary keywords. An assumption could be made that the keywords might exist in the original orders, outside order sets, or collective knowledge on which they depend. The methods used in this study made it difficult to ascertain exactly which prior orders or order-sets were being referred to.

Modification orders

Similar to referent orders, another complexity encountered was the existence of orders that modified a previous order. There exists an implication here is that a previous order exists. In the present keyword-based scheme, only the "Discontinued Orders" category contained modification orders. An example of a modification order resulting in an increase in the drug dosage is:

- "Increase Enalpril to 1.25 mg IV q 6 hrs"

When considering the implications of using a speech-recognition order entry system, it remains to be seen whether or not the knowledge needed to process an order like this can be maintained. Perhaps the user will be required to discontinue the old order and replace it with the new order.

Conditional orders

Conditional orders represent another layer of complexity not accounted for by the present keyword-based categorization scheme. Conditional orders often incorporate a specific instruction and/or condition. In some cases, the order is to be executed only when the condition is met. Some examples of these orders include:

- "Nipride drip to keep SBP < 160 > 140"
- "Call HO if Temp <35.3"

While not problematic for Level 1 of the model, this type of order may be hard to interpret in Levels 2 and 3.

Table 9 shows the breakdown of these three types of complex orders as they occurred in the Test Set. All together, these types of complexities comprise almost 25% of the Test Set.

Complex Order Type	Number of Orders	% of Total Orders in Test Set
Modification Orders	72	12.0
Referent Orders	29	5.0
Conditional Orders	45	7.2
Total	146	24.2

Table9. Number of orders in Test Set that were complex.

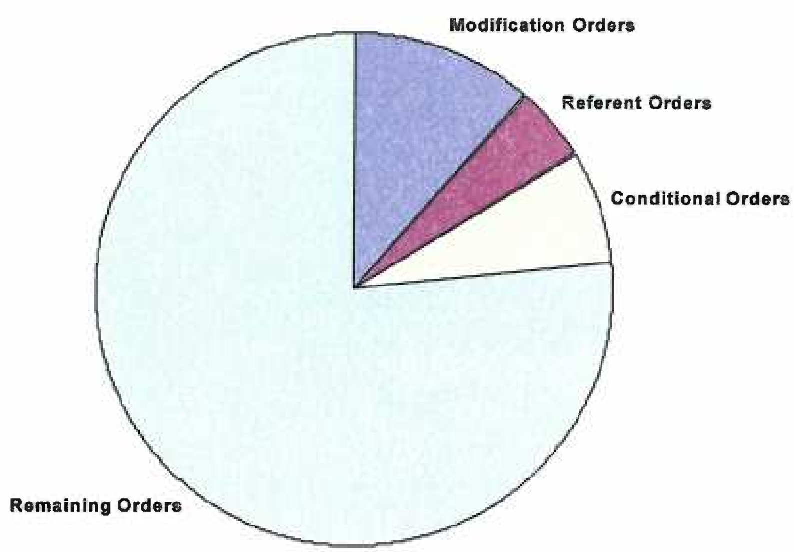


Figure8. Distribution of three different types of complex orders encountered in this study within the Test Set.*N = 625 orders.*

Limitations and Lessons Learned

It is clear that the program, for Level 1 of the model in Figure 1, was too simple to accommodate some of the complexities found in the data, in particular, the combination, modification, conditional and referent orders. In many cases, these complexities did not prevent the orders from being placed in the correct categories; however, their function of modifying, changing, or referring to other orders was not examined at this level. These details could potentially be dealt with further along in the model design (Levels 2 and 3) with future study. If the design of the program allowed for a single order to be categorized in more than one category or if there where a mechanism for appropriately parsing the distinct components of the orders, the problem with the combination orders may have been eliminated.

Categories and Keywords

During the development phase, the task of "training" the PERL script, of adjusting the keywords and categories, was a tedious one. Despite this, the major reason for incorrect categorization was a lack of keywords reflecting an inadequate development of key word list. If the PERL script had looked for combinations of keywords (i.e., forming some simple rules) rather than individual keywords, greater sensitivity might have been achieved. Additional gains might also have been achieved by adding more keywords. This could have been accomplished by utilizing a larger development set and/or greater collaboration with domain experts.

Modeling from handwritten orders

As stated in the methods, handwritten orders were chosen for this study because they were relatively easy to obtain and reflected a reasonably unconstrained use of vocabulary (as compared to pre-printed order forms, for example). One of the drawbacks, not surprisingly, was the interpretation of handwriting. Admittedly, this got easier as the researcher became familiar with the individuals' handwriting styles, terms and abbreviations, but it still made the task of interpretation and transcription a slow one.

Adding to this challenge was the fact that orders were over 2 months old when they were obtained and older still by the time they were transcribed. If there were confusing abbreviations or terms, the amount of elapsed time rendered it virtually impossible to contact the originating clinicians, especially since many of them were resident physicians who had left the OHSU campus to continue their rotations elsewhere.

These challenges may have been further heightened by the fact that the researcher only kept photocopies of the order sheets and not any other part of the medical record. Progress notes might have also provided insight into some of the more confusing orders. In future studies, some of these issues may be effectively addressed by using data collection methods that focus on a smaller group of clinicians, over a narrower time frame. This approach should allow for greater opportunity for clarification from the originating source.

All challenges aside, the handwritten orders did provide a rich data source and demonstrated much variety in vocabulary chosen to express orders. Further study is required to answer the question of whether handwritten orders provide an adequate starting point for modeling the ICU language for a future spoken language interface, because clinicians may well verbalize orders differently from the way they write them.

Abbreviations

Another problem encountered in this study was use of abbreviations. Many of the chosen keywords consisted of frequently occurring abbreviations, and their abundance in the data demonstrates their timesaving benefit when writing orders by hand. One can speculate, that in a spoken language, the use of abbreviations would probably not be as prevalent, because speaking is normally much quicker than handwriting.

Some abbreviations, however, have become so ubiquitous that they are frequently verbalized the same way. Some examples are "CBC" for "cell blood count" and "EKG" for "electrocardiogram." While these are quite familiar, for many of the abbreviations encountered, it is not as evident how they might typically be verbalized in a spoken dialogue. An example could is use of the abbreviation "mg" for the word "milligrams." Informal interviews of clinicians have indicated that this abbreviation can be verbalized both ways, for example, "Lovenox 40 *mg* subQ…" or "Lovenox 40 *milligrams* subQ…" In the handwritten data set, "milligrams" was never spelled out and was always abbreviated and hence the program searched only for the keyword "mg." If there are other examples of pronounced abbreviations, more keywords may need to be added to the model for use with a spoken interface.

Methods for Development and Testing

The methods for the development and testing of the model in this research were relatively different. The development phase (writing the program and selecting the categories and keywords) was largely conducted by the researcher with significant contributions in the form of clarification from a collaborating domain expert (an OHSU ICU clinical nurse specialist) and other clinicians (nurses, physicians and a pharmacist). There was not a process for expert consensus during the development phase as there was for the testing phase. The consensus process revealed some significant inadequacies in the researcher's conception of which orders should belong to which categories and the subsequent selection of representative keywords. Many of the issues regarding lack of keywords, particularly in the Nursing category, could have been resolved if a similar consensus process occurred during the development phase.

Implications for future research

There are several possible directions for launching future research from this work. One is to attempt the same goal as this study, developing the Level 1 of the model, but employing different methods and techniques that could result in higher sensitivity and specificity. Several ideas have already been suggested in the previous "Limitations and Lessons Learned" section above. There also remains significant unused data (photocopied order sheets) from which several more random samples could be obtained for future studies.[1]

Development of Other Levels in the Model

Another direction of research could be the development of the other levels of the language model, such as Levels 2 and 3. An example would be to develop and test domain-specific vocabularies, such as for Medications. This could include all terms for dose, route, frequency and duration. The development and testing of domain specific vocabularies could follow similar methods as this study (e.g., using a PERL script), but incorporate some of the improvements recommended above.

It is at these lower levels in the model that error checking occurs, and it becomes more important that all the components of a categorized order are identified properly. Unlike the first level, it is not sufficient to simply recognize a string of words as a "Medications" order; all the corresponding components need to be appropriately recognized. The number "3" that corresponds with "3 mg" needs to be stored as "dose" information and not as "duration." The task should be made easier when orders are already sorted into categories at the first level, but it may still be necessary to use a more sophisticated parsing program that employs some simple rules. Using the previously mentioned example, such a rule could be represented by always taking the number preceding "mg" (3, in this case) and storing it as part of "dose" information.

[1] All photocopied order sheets obtained for this study are the property of Dr. Judith Logan at OHSU.

Error checking represents yet another direction for future study and would be particularly critical in a spoken order entry system. This error checking could occur in more than one way. Voice recognition software may be unable to recognize a spoken word or may misrecognize words. One way of avoiding misrecognition is to give visual feedback to the user (i.e., showing the verbalized text on the screen) so that the clinician is assured that the spoken order was interpreted correctly by the computer. This raises an important question of how much feedback is acceptable in the ICU environment? Considering the high pace of an ICU, feedback that is slow and tedious is unlikely to be tolerated. Future studies could try to identify the threshold of feedback that is acceptable on the one hand, while still effective in allowing users to catch errors.

Another type of error checking comes from within the computer in the form of decision support rules. Use of decision support could be explored, on a domain-specific basis. A simple version of this could be a program that looks for missing components in a medication order, such as "Length of Treatment," and prompts the user to supply the missing information.

One of the final pieces of the model worth studying relates to the routing of orders to the appropriate place, presumably once the components of the order have been correctly interpreted and checked for errors. An interesting study might look at how this routing occurs and all the important staff/departments who need to be involved/notified regarding a particular order. This could be examined on a domain-specific basis, looking at just medications, or on a broader scale examining all types of orders.

The ICU Language

Another area worth studying is the vocabulary used in ICU orders and their typical pronunciations if any such pronunciations can be deemed typical. When considering future speech applications, some of the chosen keywords may not work in a spoken order-entry system. While their handwritten counterparts may be distinct, their vocal pronunciations might be so similar that it could make speech recognition difficult. For example, the commonly pronounced abbreviations, "CT" (computerized tomography) and "PT" (prothrombin time) have similar sounding pronunciations, as do "tid" (three times a day) and "bid" (twice a day). A study could look at how frequently such "sound-alike" expressions occur in a random sample of orders.

To answer the question of whether or not handwritten orders provide a satisfactory starting point upon which to build a language model, future study could test the model developed from this research on spoken orders obtained by audio recording. The results from running typed transcripts from these recordings through the same program should indicate whether or not there is significant difference between handwritten orders and spoken orders

Integration into Workflow of an ICU

While this study provides some foundation that could serve to aid in the development of spoken order-entry interfaces, it still leaves questions as to whether a spoken language interface could actually be incorporated into the work routine of an ICU. Despite a tempting "Wouldn't it be neat?" factor, would such a tool really be used during a hands-occupied task in a noisy environment, with competing communications going on between clinicians? The competeing-noise component of this question could be examined by a study that involved having a researcher (and/or some clinicians) repeat basic phrases using a commercial speech recognition product in an ICU environment. These utterances could be recorded, over various shifts, and measured to see how many of those utterances were incorrectly recognized due to noise interference. Arguably, the most may be learned in this area by testing a prototype speech-recognition order-entry system in that environment.

CONCLUSION

This thesis sought to explore three related questions. The first was to validate or invalidate the notion that has steered manufacturers of speech recognition systems away from intensive care settings. *Is the ICU language and work environment too complex for speech recognition systems?*

While this study didn't address the first question in a global sense by looking at all aspects of documentation that occur in ICUs, it did examine one piece of the documentation task: medical orders. More specifically, this study tested a scheme for categorizing medical orders and asked, secondly, *whether or not a limited vocabulary of keywords could be used to correctly categorize a random sample of ICU orders.*

The overall sensitivity and specificity of the keyword-based categorization scheme used in this study was probably not high enough to be acceptable in an ICU. However, it should be emphasized that this was a first pass at testing Level 1 of the language model and, as such, started with a relatively simple scheme, looking at vocabulary alone. When this is kept in mind, the results appear much more promising, especially in such categories such as Medications / Fluids / Transfusions, Discontinued Orders, Imaging, Diet and Laboratory tests, each of which had a sensitivity > 0.9. With some of the modifications suggested in the "Discussion" section above, particularly incorporation of some simple rules into the categorization scheme, higher sensitivities and specificities seem attainable.

The third question asked was *whether or not a language model of ICU orders could be created from handwritten physician orders*. The answer to this question will wait until future studies develop other levels of the model, and especially when an actual prototype system can be tested in an ICU environment. However, the initial results from developing the first part of the model are promising and lean toward an affirmative answer. Future studies that look at audio recordings of verbal orders may shed more light on this question, however, the ease of obtaining copies of handwritten orders makes them a far more feasible data source than making audio-recordings for this initial work.

The selection of handwritten orders for this study succeeded in providing a wonderfully diverse set of orders, with similar orders written many different ways. For example, some orders were laden with abbreviations while others used few abbreviations. Some

Medication orders were written with all the expected components written out (dose, frequency, duration, etc.), while others were less complete, perhaps referring to orders previously written (or protocols), or to the collective, undocumented, knowledge maintained among the caregivers for a given patient. This diversity, it is anticipated, more closely mimics how orders are and will be verbalized in the more unstructured realm of spoken language. Ironically, while it was order diversity that the researcher was seeking, this very diversity added complexity to the categorization task.

Finally, if refined versions of the techniques applied in this research prove successful with future development, then there is potential to apply the same techniques to other medical domains, such as in the Emergency Department.

REFERENCES

1. Teel, M.M., et al., *Voice-Enabled Structured Medical Reporting.* Computer Human Interface, 1998: p. 18-23.
2. Ash, J.S., et al., *Investigating physician order entry in the field: lessons learned in a multi-center study.* Medinfo, 2001. **10**(Pt): p. 1107-11.
3. Berg, M., *Medical work and the computer-based patient record: a sociological perspective.* Methods of Information in Medicine, 1998. **37**(3): p. 294-301.
4. Grasso, M., *Automated Speech Recognition in Medical Applications,* in *M.D. Computing*. 1995. p. 16-23.
5. Makhoul, J. and R. Schwartz, *State of the art in continuous speech recognition.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(22): p. 9956-63.
6. Schwartz, L., et al., *Voice Recognition in Radiology Reporting.* American Journal of Radiology, 1997. **169**: p. 27-29.
7. Safran, C. and P. Gertman, *An Update on Speech Recognition*, in *M.D. Computing*. 1996. p. 207-209.
8. IBM, *Website for IBM ViaVoice for Windows Pro-USB edition*. 2001, IBM.
9. Sakuma, P.A., *Speech recognition apps come of age*. 2001, USA Today and Associated Press.
10. Zafar, A., J..M. Overhage, and C.J. McDonald, *Continuous Speech Recognition for Clinicians.* Journal of the American Medical Informatics Association, 1999. **6**(3): p. 195-204.
11. Ramaswamy, M.R., et al., *Continuous speech recognition in MR imaging reporting: advantages, disadvantages, and impact.* AJR Am J Roentgenol, 2000. **174**(3): p. 617-22.
12. Wulfman, C., et al., *Graphical access to medical expert systems: V. Integration with Continuous speech -recognition.* Methods of Information in Medicine, 1993. **32**: p. 33-46.

13. IBM, *Taking dictation from doctors: IBM's MedSpeak/Radiology technology converts comments about patient X-rays into written reports.* 1997, IBM Corporation.

14. Lai, J. and J. Vergo. *MedSpeak: Report Creation with Continuous Speech Recognition.* in *Conference Proceedings on Human Factors in Computing Systems.* 1997. Atlanta, GA, USA.

15. Molnar, B., et al., *Development of a Speech-Based Dialogue System for Report Dictation and Machine Control in the Endoscopic Laboratory.* Endoscopy, 2000. **32**(1): p. 58-61.

16. Grasso, M., D. Ebert, and T. Finin. *Acceptance of a Speech Interface for Biomedical Data Collection.* in *AMIA Annual Fall Symposium, Proceedings.* 1997. Philadelphia: Hanley & Belfus, Inc.

17. Shiffman, S., et al., *A continuous-speech interface to a decsion-support system: I. Techniques to accomodate for misrecognized input.* J Am Med Inform Assoc, 1995. **2**(1): p. 36-45.

18. Ash, J.S., et al. *Perceptions of house officers who use physician order entry.* in *American Medical Informatics Society, Annual Fall Symposium.* 1999. Washington, D.C.

19. Isaacs, E., et al., *Graphical access to medical expert systems: IV. Experiments to determine the role of spoken input.* Methods of Information in Medicine, 1993. **32**: p. 18-32.

20. Cohen, P.R. and S.L. Oviatt, *The role of voice input for human-machine communication.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(22): p. 9921-7.

21. Monnich, G. and T. Wetter, *Requirements for speech recognition to support medical documentation.* Methods of Information in Medicine, 2000. **39**: p. 63-69.

22. Gosbee, J.W. and M. Clay. *Human factors problem analysis of a voice-recognition computer-based medical record.* in *Sixth Annual IEEE Symposium on Computer-Based Medical Systems.* 1993. Ann Arbor, MI, USA: IEEE Computer Society Press.

23. Bennett, D. and J. Bion, *ABC of intensive care: organisation of intensive care [see comments]*. Bmj, 1999. **318**(7196): p. 1468-70.
24. Barie, P.S., M.D. Bacchetta, and S.R. Eachempati, *The contemporary surgical intensive care unit. Structure, staffing, and issues.* Surg Clin North Am, 2000. **80**(3): p. 791-804, vii.
25. Baggs, J.G., et al., *Association between nurse-physician collaboration and patient outcomes in three intensive care units [see comments]*. Crit Care Med, 1999. **27**(9): p. 1991-8.
26. Gorman, P.N., *Personal Communication*, L. Fournier, Editor. 2000: Portland.
27. Lovis, C., et al., *Evaluation of a command-line parser-based order entry pathway for the Department of Veterans Affairs electronic patient record.* J Am Med Inform Assoc, 2001. **8**(5): p. 486-98.
28. Boyce, S., *Natural Spoken Dialogue Systems for Telephony Applications*, in *Communications of the ACM.* 2000. p. 29-34.
29. Johnson, K., et al., *A history-taking system that uses continuous speech recognition.* Proc Annu Symp Comput Appl Med Care, 1992: p. 757-61.
30. Lai, J., *Personal communication*,, L. Fournier, Editor. 2000: Portland.
31. Imhoff, M., *Acquisition of ICU data: concepts and demands.* Int J Clin Monit Comput, 1992. **9**(4): p. 229-37.
32. Hripcsak, G. and A. Wilcox, *Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance.* J Am Med Inform Assoc, 2002. **9**(1): p. 1-15.
33. Schwartz, R.L. and T. Christiansen, *Learning Perl*. 10 ed, ed. S. Talbott. 1997, Sebastopol, CA: O'Reilly & Associates, Inc. 271.
34. Davis, N.M., *Medical Abbreviations: 15, 000 Conveniences at the Expense of Communications and Safety.* 10 ed. 2001, Huntingdon Valley, PA: Neil M Davis Associates. 426.
35. Hart, J.t., R. Collier, and A. Cohen, *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge studies in speech science and communication. 1990, Cambridge [England]; New York: Cambridge University Press. 212.

APPENDIX A

Order Selection and Labeling: "Pt" = code assigned to set of orders for a patient. "Selected" = included in the study; "Order # " = unique number assigned to an order.

Pt	# of pages of orders	Selected (Y = Yes)	Selected Pages	Total Orders	Order #	Notes
A	9	Y	1,2,3	38	1-38	Only pages 1,2,3 are ICU
B	7	Y	1,3,4,5,6	54	39-92	Void #27 – not an order
C	3	N				
D	9	Y	3,4,5,6,8	42	93-134	Void #116 – not an order
E	2	Y	1,2	26	135-160	Void #160 – not an order
F	39	N				
G	28	Y	2,6,16,19,22	42	161-202	Void #189, 197 & 202 – not orders
H	34	Y	2,10,12,15,26	32	203-234	
I	18	Y	8,11,16,17,18	29	289-317	
J	18	Y	1,3,8,14,17	53	235-287	Out of order. #288 – doesn't exist
K	18	N				
L	4	Y	1,2,3,4	27	318-344	
M	18	N				
N	7	Y	1,2,3,4,7	34	345-378	
O	5	Y	1,2,3,4,5	38	379-416	
P	12	Y	1,3,4,7,10	53	417-469	
Q	115	Y	12,43,56,62,86	37	470-506	
R	127	N				
S	24	N				
T	5	Y	1,2,3,4,5	47	507-553	
U	4	Y	1,2	21	554-574	Void Pages 3-4, they were same as patient E
V	10	Y	1,2,3,4	38	575-612	Pages 5-10 were not ICU
W	68	N				
X	7	Y	1,3,4,6,7	46	613-658	
Y	4	Y	1,2,3,4	35	659-693	
Z	11	Y	1,5,8,10,11	60	694-753	
AA	6	N				
BB	25	Y	3,5,8,19,24	34	754-787	
CC	8	Y	1,2,4,5,7	31	788-818	
DD	5	N				
EE	19	Y	4,8,12,15,17	39	819-857	
FF	5	Y	1,2,3,4,5	22	858-879	
GG	7	Y	1,3,5,6,7	46	880-925	
HH	1	N				
II	38	N				
JJ	2	N				Possibly not ICU
KK	13	Y	1,3,4,6,10	34	926-959	
LL	21	N				
MM	24	Y	5,6,8,12,23	21	960-980	
OO	15	Y	2,4,10,11,14	40	981-1020	
PP	2	N				
QQ	2	N				
RR	33	N				
SS	14	N				
TT	3	N				
UU	3	Y	1,2,3	16	1021-1036	
VV	6	N				
WW	18	Y	2,6,7,12,17	40	1037-	Page 1 was a pre-printed form

					1076	
XX	4	Y	1,2,3,4	29	1077-1105	
YY	4	N				
ZZ	4	N				
AB	2	Y	1,2	25	1106-1130	
AC	7	Y	1,3,4,6,7	47	1131-1177	Page 2 was a pre-printed form
AD	7	N				
AE	6	Y	1,3,4,5,6	43	1178-1220	Page 2 was a pre-printed form
AF	1	Y	1	17	1221-1237	
Totals	911			1230		**Minus 6 for voided orders

APPENDIX B

Rules for limited modifications of orders-sheet entries during transcription to allow the PERL script to function properly.

IF	→	**THEN**
°	=	hr or hrs or degrees
$\bar{c}$	=	with
$\bar{s}$	=	without
$\bar{a}$	=	before
$\bar{p}$	=	after
✓	=	check
re - ✓	=	recheck
↓	=	**decrease** / reduce / lower
↑	=	**increase** / raise / elevate
∅	=	none, nothing, no
$\bar{X}$	=	except
$\dot{\mathrm{T}}$	=	Once or **1**
$\ddot{\mathrm{TT}}$ =	=	**2** or twice
△	=	change
DX:, Diag:, Diagnosis …		Dx:
I&O, I+O, I/O…		IO's
Cond:, …		Condition:
Act:, ACT:…		Activity:
All, Aller…		Allergies:
Tx, Trans., …		Transfer
U…		units
O_2 …		O2
S/p, SP,…		s/p
foley		Foley
I/V, I.V….		IV
®, …		(R) , (L)…
sq, subQ, …		SQ

3mg, 5%, 1x qd, … and units)	3 mg , 5 % , 1 x q d , (**add spaces** btwn numbers
DC,dc, D/C…	D/C (unless context is Discharge – then Discharge)
PT (physical therapy)…	P.T.
RT (respiratory therapy)…	R.T.
OT (occupational therapy) …	O.T.
Po, PO, …	PO
Npo, npo…	NPO
PT (prothrombin time)…	PT
PTT (partial prothrombin..)	PTT
Mcg, MCG…	mcg
Mg, mg..(magnesium)…	Mg
Mg, mg (Milligram)…	mg
Meq, MEq…	meq
Mkm, mkm…	mcg/kg/min
R/O, ro…	r/o
BID, Bid…	bid
ICA, Ica…	ICa
GTT, Gtt	gtt
AM, Am, am…	am
PRN, prn, Prn…	prn

APPENDIX C

Glossary of abbreviations encountered in both the test and development order sets.

AAOC = Antacid of choice
ABG = Arterial blood gas
ACT = Activity or Activated Clotting Time
ACW = Anterior chest wall
ADAT = Advance Diet As Tolerated
AKA = Above Knee Amputation
ALL = Allergies
Alb = Albumin
alt = alternate
amps = ampules
ANV = Acute Nausea and Vomitting
BG = Blood Glucose or Blood Gas
BID = Twice a day
BM = Bowel movement
BMS = Basic Metabolic Screen
CAD = Coronary Artery Disease
C&S = Culture and Sensitivity
c/o = complains of
CBC = Cell blood count
CBG = Capillary blood glucose
Chem. = Blood chemistry test
CHF = Congestive heart failure
CI = Cardiac Index
CMS = comprehensive metabolic screen
CNS = Central nervous system
Cond. = Condition
CPAP = Continuous positive airway pressure
Cryo = cryosurgery, cryoablation
CSF = Cerebral Spinal Fluid
CT = Computerized Tomography
CVC = Central Venous Catheter
CVP = Central Venous Pressure
CW = Chest Wall
Cx = Culture
CXR = Chest x-ray
D5 = Dextrose 5 %
DB&C = Deep breathing and coughing
D/C = Discontinue
DC = Discontinue or Discharge
DDAVP = Desmopressin acetate

DHT = Dobhoff (feeding) tube
DKA = Diabetic ketoacidosis
DNAR = Do not attempt resuscitation
DNR = Do not resuscitate
ECASA = Enteric coated aspirin
ECG = Electrocardiogram
EEG = Electroencephalogram
EKG = Electrocardiogram
Epi = Epinephrine
ES = Extra strength
ESLD = End-stage Liver Disease
ET = Endotracheal Tube
EVD = External ventricular vein
FFP = Fresh Frozen Plasma
FiO2 = Fraction of inspired oxygen
FMV = Fluorouracil
FT = Feeding tube
FTG = Foley to gravity
Fx = Fracture
GI = Gastrointestinal
GIB = Gastrointestinal Bleeding
gtt = drops
HCTZ = Hydrochlorothiazide
HOB = Head of bed
HS = Every night, bedtime, nightly
ICH = Intracranial hemorrhage
IHSS = Ideopathic Hypertropic Subaortic Stenosis
IJ = Internal jugular
IS = Incentive spirometer
IVF = IV fluids
IVP = IV Push
JP = Jackson Pratt (drain)
LOC = Laxative of choice
LP = Lumbar puncture
LR = Lactated Ringers (solution)
MAP = Mean arterial pressure
MCA = Middle cerebral aneurysm
Mcg = Micrograms
MD = Midnight
MDI = Metered dose inhaler
mEq = milliequivalents
mmol = millimoles
MOM = Milk of Magnesia
MR = may repeat
MS = Morphine sulfate or mental status
MSO4 = Morphine sulfate

MVD = Mitral valve disease
NC = Nasal cannula
Nebs = Nebulizer (handheld)
Neo = Neosysnephrine (aka phenylephrine)
NG = Nasogastric
NGT = Nasogastric tube
NKDA = No known drug allergies
Noc = Nocturnal or Nightly
NPO = Nothing by Mouth
NQWMI = Non-Q-wave myocardial infarction
NS = Normal Saline
NSG = Nuerosurg
NT = Nasal tracheal
NTG = Nitroglycerin
NTP = Nitropaste
OOB = out of bed
Ox = Oximetry
P+T = peak and trough
PB = Barometric pressure or Piggy-back
PC = Productive cough
PCN = Penicillin
pCXR = Portable Chest x-ray
Phos. = Phosphorus test
PEEP = Positive end expiratory pressure
pft = per feeding tube
PICC = Peripherally inserted central catheter
PO = by mouth, orally
PR = per rectum
PRBC = Packed Red Blood Cells
PRN = as needed
PS = Pressure Support
PT/OT = Physicial therapy/ occupational therapy
q = every
QID = four times daily
R/O = rule out
Reg = Regular (Insulin)
RR = Respiratory Rate
R/S = Restart, resume or reschedule
RT = Routine
S+S = Swish and swallow
S/P = status post
SAH = Sub-arachnoid hemorrhage
SaO2 = Arterial oxygen percent saturation
Sats or sat = saturation
SBP = Systolic Blood Pressure
SCA = superior cerebral artery

SCD's = Sequential Compression Devices
SIMV = Synchronized intermittent mandatory ventilation
SQ = subcutaneously
S&S = Swish and Swallow
supp = suppository
SW = Social Work
T ° = temperature
T&C = Type and Cross
TIPS = Transjugular intrahepatic portosystemic shunt
T.O. = Telephone Order
TCD's = Trans-Cranial Dopplers
TF = Tube feeding
TID = Three times a day
TKO = To keep open
TLC = triple lumen catheter
TV = Tidal Volume
Tx = Transfer or Treat
U = Unit
U/S = Ultrasound
UO = Urine output
VBG = Venous blood gas
V.O. = Verbal Order

APPENDIX D

List of keywords, grouped by category, used in the PERL script to identify order types.

CATEGORIES (BOLD) and Keywords	DEFINITION
DISCONTINUED ORDERS	
D/C	Discontinue
Discontinue	
Cancel, cancel	
MEDICATION/FLUIDS/TRANSFUSIONS	
mg	milligram
U, units, unit	
supp, suppository	suppository
meq	milliequivalent
mcg	microgram
PO, po	PO (oral, by mouth)
puffs	
mm	millimeter
mmol	millimole
amp, amps	
packet, packets	
gtt	drip
cream	
swabs	
spray	
nanogram	
Kg	Kilogram
MDI	Metered Dose Inhaler
Liters, liters, L	Liters
tab, tabs	
PCA	PCA (patient controlled analgesic)
IV, IV:	IV
SQ	SubQ (subcutaneous)
NTG	Nitroglycerin
NTP	Nitropaste
Nitropaste, nitropaste	
Epidural, epidural	
AAOC	Antacid of choice
LOC	Laxative of choice
D5	D5 (Dextrose 5%)
NS	Normal Saline
LR	Lactated Ringers
Saline, saline	
D50	D50 (Dextrose 50%)
Transfuse, tranfuse	
PRBC, PRBCs	packed red blood cells
FFP	fresh frozen plasma
Infusion, infusion	
IMAGING	
CT	CT (Computerized tomography)
CXR	Chest x-ray
pCXR	Portable chest x-ray
TCD's	TCD's (trans-cranial dopplers)
EKG	EKG (Electrocardiogram)
ECG	ECG (Electrocariogram)
EEG	EEG (Electroencephalogram)

MRI	MRI (Magnetic resonance imaging)
X-ray	
Contrast, contrast	
U/S	Ultrasound
Ultrasound, ultrasound	

LABS

Labs, Labs:, labs	
CK	CK (creatinine kinase)
CPK	CPK (creatinine phosphokinase)
CBC	CBC (cell blood count)
BMS	Basic Metabolic Screen
PT	PT
PTT	PTT (partial prothrombin time)
INR	INR (international normalized ratio)
CBG	CBG (capillary blood gas)
ABG	ABG (arterial blood gas)
BG	Blood gas
UA	UA (urinalysis)
Troponin	
Cx, Cx's, Culture, Cultures, cultures	Culture
C&S	Culture and Sensitivity
Mg	Magnesium
PO4	Phosphate
ICa	Ionized calcium
Ionized, ionized	Ionized calcium
Ca	Calcium
Na	Sodium
P+T	Peak and trough
peak	
trough	
Chem, chem	Chem (chemistry)
electrolytes	
Lipase, lipase	
K	Potassium
serum, Serum	
TSH	TSH (thyriod stimulating hormone)
Digoxin, digoxin	
Dig	Digoxin
Osmolality	
CSF	CSF (cerebral spinal fluid)
stain	
T&C	Type and Cross
Diff, diff	diff (differential)
differential	
Valproic	
Phenobarb	
analysis	
triglyceride, triglycerides	
HCT	Hematocrit
Panel, panel	
Protein, protein	

NURSING ASSESSMENTS/ACTIVITY/EQUIPMENT/RESPIRATORY CARE

Vitals, Vitals:, vitals	
IO's	IO's (Intakes and outputs)
UO	Urine output
Temp, Tm, T, temp	Temp (temperature)
HR	heart rate
SBP	systolic blood pressure
BP	BP or Blood pressure
DBP	diastolic blood pressure

pulse, pulses	
CVP	CVP or Central Venous Pressure
MAP	MAP or Mean arterial pressure
Nursing:	
drain, Drain	
JP	Jackson Pratt {drain}
Foley, foley	
Bedside, Bedside:, bedside:, bedside	
Pump, pump	
SCDs, SCD's	SCDs (sequential compression devices)
packs	
Clamp, clamp	
line	
extubate, Extubate	
Intubate, intubate	
DNR	DNR (do not resuscitate)
DNAR	Do not resuscitate
resuscitate	
PICC	PICC (peripherally inserted central catheter)
suction	
OOB	Out of bed
HOB	Head of bed
Bedrest	
Activity:, Activities, Activities:	
Flat, flat	
Flolan, flolan	
Trapeze, trapeze	
Bed, bed	
ventilator, Ventilator	
Sat	Sat (saturation)
IS	IS or Incentive spirometer
SaO2	Percent saturation (arterial oxygen percent saturation)
O2	O2 or Oxygen
vent, Vent	Vent (ventilator)
PEEP	PEEP (positive end expiratory pressure)
SIMV	SIMV (synchronized intermittant mandatory ventilation)
trach	trach (trachial)
Bronchodilators, bronchodilators	
PC	productive cough
RR	respiratory rate

DIET

TF, TF's, TFs	Tube feeding
Diet:, diet, Diet	
Dietary	
NPO	NPO (nothing by mouth)
Probalance	
feeding	
feeds, feed	
DHT	Dobhoff tube
residuals	
sips	
H2O	Water
ADAT	Advance diet as tolerated
chips	
cubes	

ADMIT/TRANSFER/DISCHARGE

Admit, Admit:	
Transfer, transfer, Tx	
Discharge, discharge	
SICU	SICU (surgical intensive care unit)

OR	OR (operating room)
MICU	MICU (medical intensive care unit)
NSU	Neurosurg (Neurosurgery)

CONSULTS

RT, R.T.	Respiratory therapy
evaluation	
Consult, consult, Consult:	
P.T.	physical therapy
Social, social	
O.T.	occupational therapy
Pysch	pyschology

DIAGNOSIS/CONDITION/ALLERGIES

Allergies:, allergy, Allergy, Allergies	
Allergic, allergic	
Diagnosis:, Diagnosis, Dx:, Dx	
Condition:, Condition, condition	
s/p	status post
Stable, stable	
Guarded, guarded	
Fair, fair	
CAD	coronary artery disease
DKA	diabetic ketoacidosis
MI	myocardial infarction

Total Keywords

185