



Toward an Automated Method for Assessing Jargon Spoken During Medical Encounters

Katherine A. King, BS, BA; Clifford A. Coleman, MD, MPH
Center for Ethics in Health Care, Oregon Health & Science University

Introduction

Background

Clinicians use medical jargon during conversations with patients,^{1,2} contributing to patient misunderstanding,³ and dissatisfaction.⁴ Teaching health professionals to replace jargon with plain language requires the ability to reliably detect spoken jargon. This typically involves cost- and labor-intensive subjective assessments by an observer. Moreover, in order to evaluate the effects of curricula designed to reduce the use of jargon, such observations must be made over large numbers of learners. This limits educators' ability to develop and evaluate anti-jargon curricula and programs.

Objective

Develop a valid and reliable automated method for detecting spoken jargon, that is relatively fast, inexpensive, and accessible.

Hypothesis

Video-recorded clinical encounters can be converted into written transcripts using automated transcription software. These transcripts can then be analyzed for jargon content using commercially-available jargon detection software. This method will have reasonable accuracy compared to gold standard manual analyses.

Methods

Participants and Design

Sixteen video-recorded Observed Structured Clinical Examination (OSCE) encounters between medical students and standardized patient actors.

Transcription

Each video-recorded encounter was transcribed by a professional transcriptionist (manual transcript), and again using a free online video management platform (automated transcript), which is not named here to avoid commercial bias. Two versions of both manual and automated transcripts were produced, one with student-plus-patient dialogue; one with student-only dialogue.

Jargon Detection

To establish a gold standard, two researchers independently reviewed each of the 16 manual transcripts, and identified jargon based on a previously published taxonomy,^{2,5} including technical terms (e.g., "hypertension"), acronyms and abbreviations (e.g., "CBC"), medical vernacular (e.g., "acute"), and unnecessary synonyms (e.g., "suppress"). We did not count medicalized English (terms with both common medical and nonmedical meanings; e.g., "negative") because we did not expect the automated jargon software to reliably detect such terms. The proprietary automated jargon detection software (not named here to avoid commercial bias) included two different jargon filters – "health words" and "non-health words." We analyzed transcripts using the "health words" filter alone, then with both filters.

Analyses

We compared "jargon rates" (number of jargon terms per thousand words) for total jargon and unique jargon terms, using two-tailed t-tests. We calculated sensitivities and specificities for various combinations of transcript type, and jargon detection filter used, as compared to the gold standard.

Results

Gold standard: Using manual transcription and manual jargon counting, the average rate for total jargon used by students was 26 per 1000 words spoken. Eliminating duplicate terms, lowered this to 18 per 1000 (**Figure 1, bars f & k**). **Automated transcription:** Compared to manual transcription, the average automated student-plus-patient word count was 9% smaller, however, the average student-only word count was not significantly different between the modes of transcription (**Table 1**).

Table 1. Mean word count by manually and automatically transcribed methods.

	Manual	Automated	p
Mean student-plus-patient word count	1860	1691	0.029
Mean student-only word count	1280	1215	0.25

On student-plus-patient transcripts, there was no difference in jargon detection software performance between manual and automated transcription for each filter type (**Figure 1, bar b vs. c, & d vs. e**); on student-only transcripts, the jargon detection software was more accurate (fewer false positives) on automated transcripts when the combined filter was used (**Figure 1, bar j vs. i**). **Automated jargon detection software:** The "health words" filter had low sensitivity, missing about 50% of jargon across all transcript types (**Figure 1, bar b vs. a, g vs. f, & l vs. k**). In contrast, the combined filter had much higher sensitivity, but also produced large numbers of false positives (**Figure 1, bar d vs. a, i vs. f, & n vs. k**). However, because total word counts were large, specificity remained high for the combined filter (**Table 2**). **Overall automated method:** applying the combined filter of the automated jargon detection software to the 16 automated transcripts yielded 288 true positives, 17953 true negatives, 712 false positives, and 254 false negatives, for a sensitivity of 53.1%, specificity of 96.2%, and overall accuracy of 95.0% (**Table 2**).

Figure 1. Jargon detection in medical encounters by manual and automated methods.

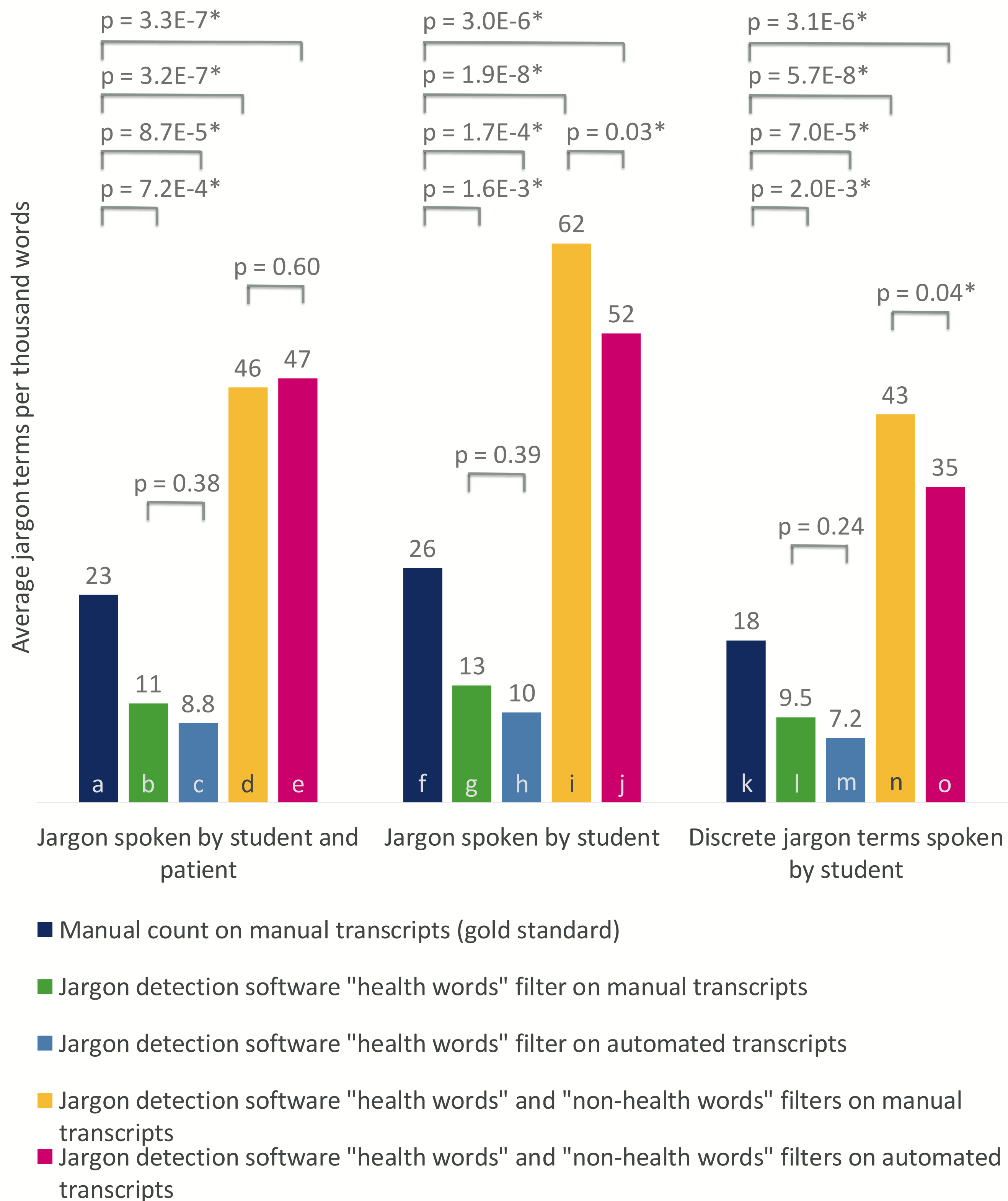


Table 2. Sensitivity and specificity of automated jargon detection software, student-only dialogue

	Manual transcripts by filter type		Automated transcripts by filter type	
	"Health Words"	"Health Words" + "Non-Health Words"	"Health Words"	"Health Words" + "Non-Health Words"
Sensitivity	45.8%	75.8%	25.8%	53.1%
Specificity	99.7%	95.9%	99.7%	96.2%
Accuracy	-	95.4%	-	95.0%

Discussion

Significance

This is the first known attempt to develop an automated method for detecting jargon in spoken medical encounters. The automated jargon detection software (which was designed for use with print materials) performed relatively poorly in terms of sensitivity, when applied to transcripts of spoken dialogue. This reflects differences in what we considered jargon, compared to the limited corpus of jargon terms used by the software's algorithm. Overall, our automated method showed marginal utility, with the sum of sensitivity (0.531) and specificity (0.962) approaching the suggested threshold of 1.5,⁶ and an overall accuracy of 95%. This study offers "proof of concept" that an automated method can provide a relatively low-cost, easy-to-use, and accessible means for assessing jargon in clinical encounters. The method may be most useful for comparing jargon rates between cohorts of learners, such as before and after a curriculum intervention.

Limitations and Considerations

While a taxonomy has been proposed,⁵ there is no consensus on what constitutes medical jargon. The terms identified in our "gold standard" assessment might differ from those of others. For example, we did not count medicalized English⁵ terms (thus, our gold standard jargon rate is likely an underestimate). This poses problems for interrater reliability, that may be alleviated through our automated approach, which removes the element of subjectivity. The automated transcripts undercounted words by 9%, apparently due to undercounts of patient dialogue (**Table 1**). We calculated sensitivities and specificities based on total rather than unique jargon rates, and for student-only dialogue (**Table 2**); a fully automated method will require analyzing student-plus-patient dialogue, which may reduce specificity.

Conclusions

- The current manual method for assessing jargon used during clinical encounters is cost- and labor-prohibitive; its subjective nature requires specialized observer training.
- Health professions educators need a low-cost easy-to-use method for identifying jargon used during spoken encounters.
- Our study provides "proof of concept" and validation of an automated approach to detecting jargon.
- While our automated method undercounts jargon by 47%, we believe it can be used for tracking individuals' jargon use, and average jargon rates between cohorts of learners, to help evaluate anti-jargon curricular interventions.
- Future efforts should expand the corpus of jargon terms recognized by jargon detection software, to improve the sensitivity of an automated approach.

References

- Castro CM, Wilson C, Wang F, Schillinger D. Babel Babble: Physicians' use of unclarified medical jargon with patients. *Am J Health Behav* 2007;31(Suppl 1):S85-S95
- Miller AN, Bharathan A, Duvuuri VNS, et al. Use of seven types of medical jargon by male and female primary care providers at a university health center. *Patient Educ Couns*. 2021;105(5):1261-7.
- O'Connell RL, Hartridge-Lambert SK, Din N, St John ER, Hitchins C, Johnson T. Patients' understanding of medical terminology used in the breast clinic. *Breast*. 2013 Oct;22(5):836-8.
- Allen KA, Charpentier V, Hendrickson MA, et al. Jargon Be Gone – Patient Preference in Doctor Communication. *Journal of Patient Experience*. 2023;10. doi:10.1177/23743735231158942
- Pitt MB, Hendrickson MA. Eradicating Jargon-Oblivion-A Proposed Classification System of Medical Jargon. *J Gen Intern Med*. 2020;35(6):1861-4.
- Power M, Fell G, Wright M. Principles of high-quality, high value testing. *BMJ Evidence-Based Medicine*. 2013;18(1):5-10

Acknowledgments: Thank you to Samantha Birk, MS for help anonymizing data, and to the OHSU Center for Ethics in Health Care, and OHSU Scholarly Projects Program for funding support.