

# Representing Normative Arguments in Genetic Counseling

Nancy Green

Department of Mathematical Sciences  
University of North Carolina at Greensboro  
Greensboro, NC 27402 USA  
nlgreen@uncg.edu

## Abstract

Our research is on developing artificial intelligence-based approaches to help lay audiences to understand medical and other scientific arguments. Currently, we are developing a natural language generation system that will synthesize patient-tailored genetic counseling documents. In this position paper, we focus on two forms of internal representation to be used by the system: a qualitative causal probabilistic model for domain reasoning, and the representation of normative arguments, based upon the Toulmin model, in document planning.

## Introduction

As the role of genetics in medicine grows, there is an increasing need for healthcare consumers to be able to make informed decisions using information on genetic conditions, genetic testing, and inheritance risks. Currently in the USA, genetic counselors meet face-to-face with clients of genetics clinics to help them understand this type of information. After the meeting, the counselor writes the client a *patient letter*, a tailored document summarizing services and information provided to the client (Baker et al. 2002). We are designing a prototype natural language generation (NLG) system (Hüske-Kraus 2003; Cawsey et al. 1997) that will generate the first draft of a patient letter to assist genetic counselors. Using a graphical interface designed to minimize typing, the counselor will provide the system with information such as the patient's symptoms, pedigree, and test results and the clinic's diagnosis. Note that our research focus is on NLG rather than decision-support tasks such as genetic diagnosis or calculation of genetic risk. It is assumed that those tasks will be performed by experts outside of our system. Given the above information about a case, our system will use domain knowledge to infer the basis for the experts' conclusions provided by the counselor, and then use

communication knowledge to synthesize the first draft of a letter.

We began by studying a corpus of patient letters written by genetic counselors. The letters contain arguments serving medical documentation, legal, and educational goals. For example, an argument may be given to document the clinic's reasoning in case support for a diagnosis changes as new evidence is acquired; an argument may be included for legal purposes to justify which tests were performed; or an argument may be given to help the client understand the genetic basis for her child's condition and the risk to her future offspring. One of our working assumptions is that the NLG system must be informed by knowledge of what constitutes a normative argument in clinical genetics.

In addition to studying use of arguments in the corpus, we analyzed the general types of biomedical concepts and their causal interrelationships communicated in the letters and identified a set of categories (e.g., symptom, test result, genotype) with good inter-rater reliability (Green 2005). The motivation for that study was to enable us to construct a domain model that reflects the simplified conceptual model of medical genetics communicated by genetic counselors to their lay clients. To qualitatively model the types of reasoning illustrated in the corpus, we implemented a prototype domain model as a qualitative causal probabilistic network (QPN) (Green, Britt, and Jirak 2004).

The NLG system will use knowledge of normative argument form to select and organize information from this domain model. In other words, while a QPN provides a useful internal representation that enables the system, e.g., to understand the relationship between a patient's symptoms and the diagnosis made by the patient's medical provider, it provides no guidance on how to present the information. On the other hand, knowledge of normative argument form is not sufficient. The system also must make use of knowledge of the general outline of a patient letter: discussion of reasons for referral (including the tentative diagnosis before testing), tests performed and their results, final diagnosis, recurrence risks, etc. Then, within each one of these genre-specific

topics, the system can construct arguments for claims to be presented on the current topic.

In addition to the above NLG tasks, the system must make a variety of rhetorical and linguistic choices that should have a significant effect on how well a lay audience can understand the arguments, e.g., determining presentation order of components of an argument and which argument components to present explicitly, choice of rhetorical cues (via word choice and syntactic form), and choice of content words and phrasing. An important issue in choice of content words and phrasing is the expression of probabilities in a manner that is comprehensible to a lay audience.

In this paper, our focus is on the relationship of the causal probabilistic domain model to normative argument structure. We shall analyze an excerpt from the corpus in terms of these two models and discuss how they are related.

### Causal Probabilistic Representation

For general information on causal probabilistic networks see, e.g., (Korb and Nicholson 2004). Figure 1 shows a (simplified) subgraph of the system's domain model representing the biomedical content of the following excerpt from a genetic counseling letter in the corpus. In the excerpt, sentences have been numbered for discussion purposes, and information in square brackets has been used in place of confidential information.

- (2) [DOCTOR] asked us to evaluate [PATIENT] to determine if [HIS/HER] delays in development and [BIRTH DEFECT] were due to a recognizable genetic condition.
- (3) During your appointment on [DATE], we obtained a blood sample from [PATIENT].
- (4) ... a special analysis ... was done to test for Velocardiofacial syndrome (VCF).
- (5) Individuals with VCF often have [BIRTH DEFECT] and learning problems.
- (6) A newly developed test ... was also done.
- (7) [PATIENT'S] ... study ... showed a normal result.
- (8) It is important to remember that, even though these tests were negative, [PATIENT'S] problems could still be caused by a genetic alteration.

The network in Figure 1 contains four nodes, each representing a discrete random variable. The domain of each variable is given in parentheses. For more information about the categories of variables that may be used in the domain model see (Green 2005). The directed arcs in this network represent direct causal dependencies between nodes. The strength of the relationship between variables is represented by means of qualitative influence relations (Green, Britt, and Jirak 2004).

For the case represented in the above excerpt, a genetic counselor using the system to draft a letter would select the initial diagnosis of VCF. The system represents this initial diagnosis in the network in Figure 1 by setting the probability to *possible*, at time *pretest*, that the genotype-type node labeled **VCF genotype** has the value *1 mutated allele*. (VCF is a condition due to having one mutated allele of a certain gene in all of an individual's cells.) The system's graphical interface then provides a list of possible symptoms of VCF and related disorders. After the genetic counselor has indicated which symptoms have been observed, a birth defect and learning problems, this would be represented in the network by setting the probability to *certain* that the value of the symptom-type node labeled **Birth defect and learning problems** is *yes*. The arc from **VCF genotype** to **Birth defect and learning problems** represents the potential causal relation between the occurrence of VCF (i.e., having one mutated allele for VCF) and those symptoms. Thus, qualitative domain reasoning from observations is consistent with the clinic's initial diagnosis.

Similarly, the genetic counselor could indicate what tests have been performed and their results. For the case discussed in the excerpt, the inputs to the system would be represented in the network by setting the probability to *certain*, at time *posttest*, that the value of the result-type node labeled **VCF test results** is *negative*. The arc from **VCF genotype** to **VCF test results** represents another potential causal relation, assuming that the intervention of performing a test on the patient has been performed. Thus, domain reasoning from the new evidence is consistent with the experts' final conclusion that it is unlikely that the patient has VCF, represented in the network by a lowered probability, at time *posttest*, that **VCF genotype** has a value of *1 mutated allele*.

In the network shown in Figure 1, another possible cause of the observed systems is represented by the genotype-type node labeled **other genotype**. (The domain of this variable is specified as *0, 1, or 2 mutated alleles* since the unspecified condition could follow an autosomal recessive inheritance pattern unlike VCF, which is autosomal dominant.) The experts' final conclusion that it is possible that the patient has some other genetic condition would be represented by setting the probability to *possible*, at time *posttest*, that **other genotype** has the value *2 mutated alleles*. Given the qualitative influence relations in this network from both **VCF genotype** and **other genotype** to **Birth defect and learning problems**, this conclusion is consistent with the increase in probability, from time *pretest* to *posttest*, that **other genotype** has the value *2 mutated alleles*.

### Argument Representation

Toulmin's model of argument form (Toulmin 1998) is applicable to many of the arguments found in the corpus.

Toulmin's model was developed to describe norms of argumentation in academic disciplines (van Eemeren and Grootendorst et al. 1996). Recently, it has been used as a framework for teaching evidence-based medicine (Jenicek and Hitchcock 2005). In Toulmin's model (as summarized in van Eemeren and Grootendorst et al. 1996), the *data* are the facts used to defend a claim. The *warrant* is the (sometimes implicit) principle that licenses the claim, given the data. In addition, a *qualifier* may be used to indicate the strength of the claim or the existence of exceptions to the warrant, and a *backing* may be used to justify the warrant; e.g., one form of backing is to give the facts on which the warrant is based.

Now we shall analyze arguments conveyed in sentences (2) through (8) of the excerpt in terms of Toulmin's model, and then give some correspondences between the Toulmin argument representation and the part of a domain model shown in Figure 1. (Although we have not distinguished the qualifier component of the arguments, each claim incorporates a qualitative indicator of uncertainty.) Argument I is reconstructed below from sentences (4), (2), and (5); it outlines the support, before the test results were known, for the claim that the patient has the VCF mutation and that it is responsible for his/her symptoms.

#### Argument I

- Claim (in 4): [PATIENT] may have the VCF mutation, which is responsible for his/her [BIRTH DEFECT] and learning problems
- Data (in 2): [PATIENT] has [BIRTH DEFECT] and learning problems
- Warrant (implicit): Presence of the VCF mutation can cause [BIRTH DEFECT] and learning problems
- Backing (in 5): The conditional probability of [BIRTH DEFECT] and learning problems, given the presence of the VCF mutation, is high

Next, we analyze (7) and (8) as the realization of a chain of three arguments supporting the final diagnosis, i.e., supporting the claim of Argument IV.

#### Argument II

- Claim (implicit): Unlikely that [PATIENT] has the VCF mutation
- Data (in 7): [PATIENT'S] VCF test results were negative
- Warrant (implicit): When testing has been performed, presence of the VCF mutation should cause the VCF test results to be positive

#### Argument III

- Claim (implicit): Unlikely that the VCF mutation is responsible for [PATIENT'S] [BIRTH DEFECT] and learning problems

- Data (claim of Argument II): Unlikely that [PATIENT] has the VCF mutation
- Warrant (implicit): A precondition of *responsible* is not satisfied.

#### Argument IV

- Claim (in 8): [PATIENT] may have some other (unspecified) genetic condition, which is responsible for his/her [BIRTH DEFECT] and learning problems
- Data (claim of Argument III): Unlikely that a VCF mutation is responsible for [PATIENT'S] [BIRTH DEFECT] and learning problems
- Warrant (implicit): Either the VCF mutation or some other genetic mutation could be responsible for [BIRTH DEFECT] and learning problems

Toulmin's original model is not applicable to all of the arguments found in the corpus. For example, consider sentence (9), which follows (8) in the letter shown above.

- (9) There are currently no specific tests for many of the genetic changes that can cause birth defects and learning problems.

We analyze this as a kind of dialectical argument, i.e., an argument intended to address anticipated objections of the audience to the writer's claims, where the audience's objections are based upon incorrect information. In this example, the writer may be addressing the mistaken belief of the patient's parents that the genetic testing performed on the patient has ruled out the possibility of any genetic basis for their child's symptoms. We reconstruct this argument, which supports the same claim as the claim of Argument IV, as follows.

#### Argument V

- Claim (in 8): [PATIENT] may have some other (unspecified) genetic condition, which is responsible for his/her [BIRTH DEFECT] and learning problems
- Audience's Counter-argument (implicit):
  - Claim: [PATIENT] has no genetic condition
  - Data: [PATIENT'S] VCF test results were negative
  - Warrant: When testing has been performed, presence of any genetic condition should cause the VCF test results to be positive
- Conceded Data (in 8): [PATIENT'S] VCF test results were negative
- Counter-Warrant (implicated in 9): The VCF test does not identify other genetic conditions that can cause birth defects and learning problems.

We plan to use normative argument structures such as the above as intermediate representations during the

generation process. Given a claim, the system will extract information from its domain model to supply data (or conceded data), warrant (or counter-warrant), and backing (if any). A combination of corpus-based and empirical methods will be used to transform this internal representation into text. We assume that the transformation of domain information to text will be more straightforward if the system has access to the role of the information to be presented in the underlying arguments than if no such information were available.

Next we shall illustrate what information would be extracted from the domain model to fill the different components in some of the above arguments. For example, the part of the claim of Argument I that the patient may have the VCF mutation maps to the following in the network in Figure 1: at time *pretest*, the qualitative probability is *possible* that the variable **VCF genotype** has the value *1 mutated allele*. The other part of the claim, that this mutation is responsible for the patient's symptoms, maps to the probability of **VCF genotype** having the value *1 mutated allele* being higher than some threshold and higher than the probability of any other possible causes at time *pretest*. The data of Argument I maps to the belief that it is *certain* that **Birth defect and learning problems** is *yes*. The warrant maps to the relation of positive causal influence represented by the arc from **VCF genotype** to **Birth defect and learning problems**. The backing for the warrant of Argument I is a qualitative conditional probability statement that maps to information outside of the network justifying that causal influence relation. The mappings from Arguments II through IV to the domain model are similar to the above. The audience's counter-argument in Argument V could be mapped to a network representing a model of the audience's supposed beliefs like the one shown in Figure 1, but with an arc representing positive causal influence from **other genotype** to **VCF test results**.

## Conclusion

This paper described two forms of internal representation to be used by a natural language generation system during synthesis of the first draft of lay-oriented genetic counseling documents. The qualitative causal probabilistic model reflects the simplified conceptual model of this domain used by genetic counselors in communication with their clients. However, the domain model does not represent what constitutes a normative argument in this genre. Thus, the presentation of information from this model will be informed by a representation of normative arguments based upon the Toulmin model. The next step in our research is to focus

on the transformation of the normative argument representation into a presentation that is comprehensible to a lay audience. A combination of corpus-based and empirical methods will be used.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under CAREER Award No. 0132821.

## References

- Baker, D.L., Eash, T., Schuette, J.L., Uhlmann, W.R. 2002. Guidelines for writing letters to patients. *J Genetic Counseling* 11(5):399-418.
- Cawsey, A.J., Webber, B.L., Jones, R.B. 1997. Natural language generation in health care. *J. Am Med Inform Assoc* 4(6):473-82.
- van Eemeren, F.H., Grootendorst, R., et al. 1996. *Fundamentals of argumentation theory: a handbook of historical backgrounds and contemporary developments.*: Lawrence Erlbaum Associates, Mahway, New Jersey.
- Green, N. 2005. A Bayesian network coding scheme for annotating biomedical information presented to genetic counseling clients. *J. Biomedical Informatics* 38:130-44.
- Green, N., Britt, T., and Jirak, K. 2004. Communication of Uncertainty in Genetic Counseling Patient Education Systems. In *AAAI 2004 Fall Symposium on Dialogue Systems for Health Communication*.
- Hüske-Kraus, D. 2003. Text generation in clinical medicine – A review. *Methods Inf Med* 42:51-60.
- Jenicek, M. and Hitchcock, D. L. 2005. *Logic and Critical Thinking in Medicine*. AMA Press.
- Korb, K., Nicholson, A.E. 2004. *Bayesian artificial intelligence*. Chapman Hall/CRC, Boca Raton, Florida.
- Toulmin, S.E. 1998. *The uses of argument*. (9<sup>th</sup> ed.) Cambridge University Press, Cambridge, England.

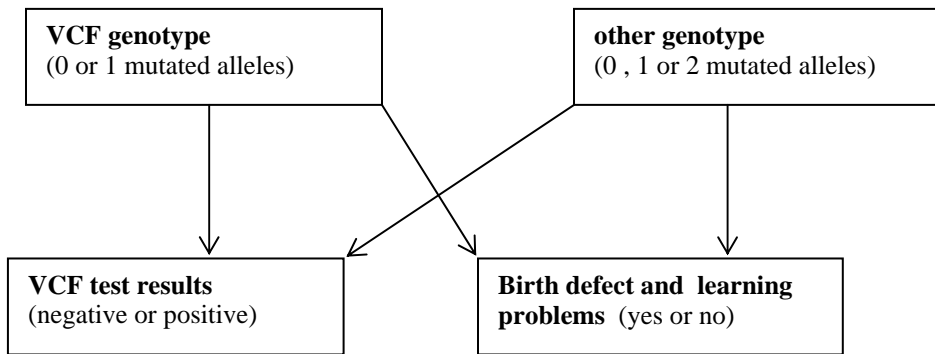


Figure 1. Causal network for excerpt.