A clinical model III: Universal diagnosis syntax.

Carl-Fredrik Bassøe

Abstract

Background

Diagnoses are crucial assets of clinical work and provide the foundation for treatment and follow up. Medical diagnoses are found by lookup in alphanumeric classifications such as ICPC-2 and ICD-10. Diagnoses should be informative and customized to the patient’s problem.

Medical diagnosis are proper names such as *Addison syndrome* and rigid descriptions like *Bacterial endocarditi*s and *Myalgic encephalopathy* (*ME*). *Bacterial endocarditi*s has a biological reference but *Myalgic encephalopathy* does not since *algic* and *encephalopathy* are heterogenous propositional attitudes. Likewise, diagnoses containing signs are propositional attitudes expressed by physicians. Propositional attitudes have no truth value and cannot be used as impartial diagnoses. Rigid and informal expressions can hide aspects of meaning and lend invalid arguments intuitive credibility. These problems cannot be overcome because the patients’ and physicians’ statements may well be normative. However, they can be minimized in two ways. First, by making hidden meanings explicit and exclude symptoms and signs from diagnoses. Second, by securing the reference(s) of diagnoses.

Results

The formula d:=e&o&p concatenates three variables e, o and p that are assigned terms that refer to a name of an etiological agent e, disorder o and pathogenetic mechanism p. The sign := assigns the character strings to a diagnosis d. An example is e:=’*Staphylococcus aureus* ‘, o:=’endocard’, p:=’it is’ that generates the diagnosis d=’*Staphylococcus aureus* endocarditis’. The formula generates diagnoses with appropriate morphology and syntax. Diagnoses formed this way are shown to comply with diagnoses from clinical practice. With certain extensions the formula generates complete, systematic medical diagnoses that are applicable to all medical specialties. The formula is extensible and versatile. It scales well with the developments in clinical medicine, systems biology, molecular biology and microbiology.

The diagnosis generating formula d:=e&o&p requires meticulous analysis of the components of diagnoses plus the introduction of appropriate terms, variables and terms. Terms partition on established clinical categories and adhere to established clinical nomenclature. Diagnoses and parts of diagnoses are terms. The syntax creates a universal diagnosis language.

Conclusions

The present study concerns a universal diagnosis syntax (UDS) that generates diagnoses using the formula d:=e&o&p with several extensions described in the study. The formula is easy to learn and covers diagnoses in all medical specialties. The present work succeeded in creating diagnoses from the formula. The fundamental insight is that no matter how complicated a diagnosis is it can be generated by a systematic process, which adds terms one by one. UDS may have implications for medical education and classifications. The formula lays a foundation for systematic and structured clinical decision-making. Formulas are hallmarks of hard science. So, d:=e&o&p anticipates a scientific clinical revolution.

Background

The introduction embraces health personnel that are not acquainted with linguistics or informatics. Use and mention of terms are distinguished typographically. Use: Lungs are in the chest. Mention: *Lungs* consists of five letters. Terms mentioned in formula are embraced by single brackets. Thus, in a formula *lung* is converted to ‘lung’. Also, *Lungs* is a string of characters in the domains [a-z,A-Z].

Current medical diagnoses are a blend of names of diseases, disorders, syndromes and clinical findings. The names of diseases, disorders and syndromes consist of proper names and rigid descriptions of clinical entities. Proper names such as *schizophrenia* and *mononucleosis* are invented and *Conn syndrome* and *Hirschprung’s disease* are surnames. Such diagnoses cannot be constructed from morphemes that refer to the etiology, disorder or pathogenetic mechanisms. For this reason, proper names do not count as diagnoses in the present sense of the word.

Systematic diagnoses such as *bacterial sinusitis*, *E. coli cystitis* and *myocardial infarction* are considered to be rigid (definite) descriptions (Gamut 1991a). All these diagnoses refer to an etiology, one or more disorders and/or pathogenetic mechanisms. In this work, only rigid descriptions count as medical diagnoses. We also require a rather stringent but informal syntax that underlies informal systematic diagnoses such as *Acute Neisseria meningitidis meningitis*, *bacterial tonsillitis*, *Streptococcal tonsillitis*, *chemical alveolitis*, *hereditary spherocytosis* and *idiopathic pulmonary fibrosis* that are obtained from standard medical textbooks and classifications.

β-cells are found in the pancreas and pituitary gland. The same name is encountered in several diagnoses. The morpheme *itis* may point to inflammation, infections, allergies and autoimmune reactions. This study aims to resolve ambiguities that arise from different meanings of terms in different contexts.

The history of medicine tells a long tale of misnomers, for example the obsolete diagnosis *pachymeningitis hemorrhagica interna* (Jones 1972). The diagnoses *hysteria* and *neurasthenia* were common about a century ago but are rarely used today. Medicine has the capacity of clearing away misnomers, but some still remain. Furthermore, diagnostic terms demand a clear morphology. For example, the term *itiscys* is an incomprehensible misnomer, whereas *cystitis* is a well-formed expression that physicians immediately understand. The diagnoses generated in this study disallows misnomers.

Clinical findings (symptoms, signs and supplementary investigations) are part of arguments used in clinical decision-making (CDM) to select and create diagnosis. For example, *dysphagia* and *dyspepsia* refer to unclear collections of clinical findings. The present work separates arguments from diagnostic conclusions (diagnoses) and does not allow collections of clinical findings as diagnoses. By doing this we can rid diagnoses of misnomers caused by mixing etiology, disorder and pathogenetic mechanisms with clinical findings.

At least 47 distinct ways for expressing *myocardial infarction* appear in clinical notes (Curé 2015). Because of the variety of such diagnoses they deserve the label natural medical language. This article concerns the development of a universal diagnosis syntax (UDS) that standardizes diagnoses and assigns only one composite form to each diagnosis.

Logic, mathematics and chemistry have profited from symbolic notations (Kenny 2000:13). Morphology and syntax fall within the domains of linguistics, logic, the philosophy of language and informatics (Robins 1997:32-3, Gamut 1991a, Gamut 1991b, Miller 2004, Jurafsky 2000). An appropriate syntax for a formal language requires a vocabulary and a set of rules (Miller 2004:6-7). Transformational-generative grammars have proved useful for language production (Chomsky 1972, Chomsky 1986, Gamut 1991a:21). Generative grammar evolved into formal language theory (FLT), which has a wide range of applications (Fitch 2012). But no such symbol system is available to clinical medicine. This study introduces symbols and a formula that generates systematic medical diagnoses.

Widely used medical classifications such as the 10th International Classification of Diseases (ICD-10) (ICD 2015) and the 2nd International Classification of Primary Care (ICPC-2) (ICPC 2015) essentially list of proper names and rigid descriptions of diagnoses. They are used in Electronic Patient Records (EPR) to select diagnoses and their associated codes. Any such classification lack important diagnoses. The lists last for some years. Extending and revising them is difficult and time-consuming, and backwards compatibility remains a serious problem.

Static classifications seem to be unsuited for the changing world of CDM. Combinatorial classifications such as the Systematic NOmenclature in MEDicine (SNOMED) may overcome these problems (SNOMED 2018). They account for novel diseases and syndromes simply by adding new elements to existing lists and allow new elements to combine with existing elements. However, SNOMED is secluded from the public and the classification lacks an underlying clinical model. Since SNOMED’s syntax has no semantics the classification cannot be validated.

Medical terminologies have advanced significantly in the last years (Cimino 2006), but how to reduce the variety of disease definitions remains an important unsolved problem (Boorse 1997, Hofmann 2010). Lack of an agreed infrastructure for terminology is identified as one of the major barriers to information interchange and integrate EPR with medical knowledge bases.

The Unified Medical Language System (UMLS) is an advanced effort towards the integration of biomedical terminologies (Humphreys 1993, McCray 1993). The Semantic Network, a component of the UMLS, is a structured description of core biomedical knowledge consisting of well-defined semantic types and relationships between them (Kashyap 2003). However, the UMLS Semantic Network manifests structural problems (Smith 2004) and can be further optimized (Frankewitsch 2004). In addition, UMLS does not provide sufficient logic-based structures (Zhang 2004). Concept-oriented and logic-based approaches are beneficial for creating categorical terminological structures.

The Generalized Architecture for Language Encyclopaedias and Nomenclatures in Medicine (GALEN) and UMLS are large thesauri (Zhang 2004). One aim of these projects is to bridge the gap between different terminology systems using a conceptual model and mapping facilities to natural language expressions and coding schemas (Carlsson 1996, Kashyap 2003, Rector 2003).

The GALEN project aims to bridge the gap between different terminology systems through a terminology server, which contains a conceptual model and mapping facilities to natural language expressions and coding schemas (Carlsson 1996, Rector 2003). Several projects have been launched to converge clinical terminologies towards a grand unified system (Rogers 1998, Spackman 1998). The complexity and high number of medical and health terminologies lead more recent projects to limit their attack on man-machine and machine-machine interoperability to limited domains (Boscá 2015, Ceusters 2015, Komenda 2015, Marc 2015, Seitinger 2016). Despite ongoing work toward shared data formats and linked identifiers, significant problems persist in semantic data integration across heterogeneous biomedical data sources (Livingston 2015). It remains difficult to establish shared identity and shared meaning.

A formal language is characterized by its vocabulary and syntax (Gamut 1991a:26). The vocabulary consists of three basic expressions: logical terms, logical variables, and auxiliary signs such as brackets. There is also a set of rules which show how expressions can be combined to make new expressions. The meaning of expressions is determined by Frege’s principle of compositionality: the meaning of a composite expression is wholly determined by the meanings of its component parts and the syntactic rule by which it was formed (Kenny 2000, Gamut 1991a). There are opponents to this view (Popper 2011:219-37) but for diagnoses we adhere to Frege’s principle.

Systematic diagnoses can be decomposed into their component parts (Rasmussen 1994). For example, *bacterial conjunctivitis* can be parsed into the etiology *bacteria*, the body part *conjunctiva* and the pathogenesis are given by *itis*. Finally, a flexible combinatorial classification, which was based on the same components was implemented in the early days of EPR in Norway (Bassøe 1983) and worked according to purpose in another EPR (Bassøe 1986, Bassøe 1988).

We tried a dynamic approach that lets physicians write ordinary textbook diagnoses into diagnosis fields in an EPR (Botsis 2010). The diagnoses were some years later associated with an ICPC-2 code. Also, physicians could change the diagnosis named associated with a code when the name was inappropriate. That this option was used shows the advantage of being able to change diagnoses.

Various clinical specialties use the same diagnoses and operate within the domains etiology, disorders and pathogenesis. This indicates that all specialties maybe based on a common formula for generating diagnoses.

This study aims to based UDS on a formula and clinically meaningful terms. The assignment of strings to variables is purely syntactic and the strings do not embody meaning by themselves (Fitch 2012). The variables of the formula are instantiated with names of concepts that are derived from an empirical clinical model (Bassøe 2007, Bassøe 2019, Bassøe 2019a). The relationship between the present syntax and its semantics is investigated separately (Bassøe 2019e).

References

Bassøe C-F, Sørli WG. EPR records and forms in primary health care. Tidsskr Nor Legeforen. 1983;103:1270-4.

Bassøe C-F. A combinatorial diagnostic system for general practice. Proceedings of the 11th Conference of the World Organisation of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians, London, 1986.

Bassøe C-F. Neutrophil functions studied by flow cytometry. In: Yen A (Ed.): Flow cytometry: Advanced Research and Clinical Applications. CRC Press, Inc. Boca Raton, Florida, 1989;95-148.

Bassøe C-F. A combinatorial diagnostic system for general practice: Evaluation of the social impact of disease by a computerized medical record. In: Hansen R, Solheim BG, O'Moore RR, Roger FH (Eds.): Lecture notes in medical informatics, Springer-Verlag, Berlin, 1988;212-4.

Bassøe C-F. The skinache syndrome. JRSM 1995a;88:565-569.

Bassøe C-F. Automated diagnoses from clinical narratives: A medical system based on computerized medical records, natural language processing and neural network technology. Neural Networks 1995a;8:313-319. http://www.sciencedirect.com/science/article/pii/089360809400076X (Accessed 2017/2/8).

Bassøe C-F. Combinatorial clinical decision-making. PhD dissertation, Department of Information Science and Media, Faculty of Social Sciences, University of Bergen, Norway, 2007. ISBN 978-82-308-0457-5.

Bassøe C-F. A clinical model. I: Health. EkviMed Clinical Informatics. 2019.

Bassøe C-F. A clinical model. II: Disorder, syndrome and disease. EkviMed Clinical Informatics. 2019a.

Bassøe C-F. A clinical model IV: The formal version. EkviMed Clinical Informatics. 2019b.

Bassøe C-F. A clinical model V: A clinical problem space. EkviMed Clinical Informatics. 2019d.

Bassøe C-F. A clinical model. IV: Medical semantics. EkviMed Clinical Informatics. 2019e.

Bassøe C-F. A combinatorial diagnostic system for general practice: Evaluation of the social impact of disease by a computerized medical record. In: Hansen R, Solheim BG, O'Moore RR, Roger FH (Eds.): Lecture notes in medical informatics, Springer-Verlag, Berlin, 1988, pp. 212-4.

Boorse C.A rebuttal on health. In: Humber JM, Almeder RF (Eds.) What is disease? Totowa: Humana Press; 1997, p. 1.

Boscá D, Maldonado JA, Moner D, Robles M. Automatic generation of computable implementation guides from clinical information models. J Biomed Inform. 2015;55:143-52. http://www.sciencedirect.com/science/article/pii/S1532046415000696 (Accessed 2017/1/17)

Botsis T, Bassøe CF, Hartvigsen G. Sixteen years of ICPC use in Norwegian primary care: looking through the facts. BMC Med Inform Decis Mak. 2010;10:11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2848129/ (Accessed 2015/10/6)

Cappello AR, Curcio R, Lappano R, Maggiolini M, Dolce V. The Physiopathological Role of the Exchangers Belonging to the SLC37 Family. Front Chem. 2018;6:122. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5913288/ (Accessed 2018/6/29)

Carlsson M, Ahlfeldt H, Thurin A, Wigertz O. Terminology support for development of sharable knowledge modules. Med Inform (Lond). 1996;21:207-14.

Ceusters W, Smith B. Biomarkers in the ontology for general medical science. Stud Health Technol Inform. 2015;210:155-9. https://www.ncbi.nlm.nih.gov/pubmed/25991121 (Accessed 2017/1/17)

Chomsky N. Language and the mind. San Diego: Harcourt Brace Jovanovich; 1972.

Chomsky N. Knowledge and language. New York: Praeger; 1986.

Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. Methods Inf Med. 2006;45 Suppl 1:124-35.

Curé OC, Maurer H, Shah NH, Le Pendu P. A formal concept analysis and semantic query expansion cooperation to refine health outcomes of interest. BMC Med Inform Decis Mak. 2015;15 Suppl 1:S8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4460622/ (Accessed 2017/1/17)

Derbyshire J. Unknown quantity. Washington: Joseph Henry Press; 2006.

Edwards N, Honemann D, Burley D, Navarro M. Refinement of the Medicare diagnosis-related groups to incorporate a measure of severity. Health Care Financ Rev. 1994;16:45-64. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4193498/ (Accessed 2015/8/1)

Fitch WT, Friederici AD. Artificial grammar learning meets formal language theory: an overview. Philos Trans R Soc Lond B Biol Sci. 2012;367:1933-55. http://rstb.royalsocietypublishing.org/content/367/1598/1933.long (Accessed 2014/5/18)

Frankewitsch T, Mueller M, Ganslandt T, Prokosch H. ICD10 (German edition) Mapping to MeSH - A Combination of Common Medical and Hidden Semantic Knowledge. Medinfo. 2004;2004(CD):1602.

Gamut LTF. Logic, language and meaning. Volume 1, Introduction to logic. University of Chicago Press, Chicago, 1991a.

Gamut LTF. Logic, language and meaning. Volume 2, Intensional logic and logical grammar. University of Chicago Press, Chicago, 1991b.

Gunderssen RB, Bassøe C-F. Clinical decision-making IV: Implementation of a clinical model. EkviMed Clinical Informatics. 2019.

ICD-10. http://apps.who.int/classifications/icd10/browse/2016/en 2015. (Accessed 2015/12/01)

ICD10 G40. http://apps.who.int/classifications/icd10/browse/2016/en#/G40 (Accessed 2017/2/8).

ICPC-2. http://www.who.int/classifications/icd/adaptations/icpc2/en/ 2015. (Accessed 2015/12/01)

Hiraoka A, Kumada T, Kudo M, Hirooka M, Tsuji K, Itobayashi E, et al. Albumin-Bilirubin (ALBI) Grade as Part of the Evidence-Based Clinical Practice Guideline for HCC of the Japan Society of Hepatology: A Comparison with the Liver Damage and Child-Pugh Classifications. Liver Cancer. 2017;6:204-215. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5473065/ (Accessed 7/12/2018).

Hofmann B. The concept of disease - vague, complex, or just indefinable? Med Health Care Philos. 2010;13:3-10.

Humphreys BL, Lindberg DA. The UMLS project: making the conceptual connection between users and the information they need. Bull Med Libr Assoc. 1993;81:170-7.

Jadaon MM. Epidemiology of Activated Protein C Resistance and Factor V Leiden Mutation in the Mediterranean Region. Mediterr J Hematol Infect Dis. 2011;3:e2011037. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3251907/ (Accessed 2018/3/20)

Jethwa A, Mink J, Macarthur C, Knights S, Fehlings T, Fehlings D. Development of the Hypertonia Assessment Tool (HAT): a discriminative tool for hypertonia in children. Dev Med Child Neurol. 2010;52:e83-7.

Johnson PJ, Berhane S, Kagebayashi C, Satomura S, Teng M, Reeves HL, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. J Clin Oncol. 2015 Feb 20; 33(6):550-8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4322258/ (Accessed 7/12/2018).

Jones FA (Ed.). Richard Asher talking sense. Pitman Medical, London, 1972, p. 32ff.

Jurafsky D, Martin JH. Speech and language processing. Prentice Hall, New Jersey, 2000.

Karnofsky Scale. http://www.anapsid.org/cnd/diagnosis/karnofsky.html (Accessed 7/12/2018).

Kashyap V. The UMLS Semantic Network and the Semantic Web. AMIA Annu Symp Proc. 2003;:351-5.

Kenny A. Frege. An introduction to the founder of modern analytic philosophy. Blackwell, Oxford, 2000.

Komenda M, Schwarz D, Švancara J, Vaitsis C, Zary N, Dušek L. Practical use of medical terminology in curriculum mapping. Comput Biol Med. 2015;63:74-82. https://www.ncbi.nlm.nih.gov/pubmed/26037030 (Accessed 2017/1/17)

Kringlen E. Diagnostikk som ideologi. [Diagnosis as an ideology]. Tidsskr Nor Legeforen. 1995;115:630-632.

Kurth J, Spieker T, Wustrow J, Strickler GJ, et. al. EBV-infected B cells in infectious mononucleosis: viral strategies for spreading in the B cell compartment and establishing latency. Immunity. 2000;13:485-95.

Livingston KM, Bada M, Baumgartner WA Jr, Hunter LE. KaBOB: ontology-based semantic integration of biomedical databases. 2015;16:126. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4448321/ (Accessed 2017/1/17)

Marc DT, Zhang R, Beattie J, Gatewood LC, Khairat SS. Indexing Publicly Available Health Data with Medical Subject Headings (MeSH): An Evaluation of Coverage. Stud Health Technol Inform. 2015;216:529-33. https://www.ncbi.nlm.nih.gov/pubmed/26262107 (Accessed 2017/1/17)

McCray AT, Aronson AR, Browne AC, Rindflesch TC, Razi A, Srinivasan S. UMLS knowledge for biomedical language processing. Bull Med Libr Assoc. 1993;81:184-94.

Miller A. Philosophy of language. Routledge, London, 2004.

NYHA. New York Heart Association (NYHA) Classification. https://manual.jointcommission.org/releases/TJC2016A/DataElem0439.html (Accessed 7/12/2018).

Popper KR. The open society and its enemies. London: Routledge; 2011.

Rasmussen J-E, Bassøe C-F: Semantic analysis of medical records. Meth Inform Meth 1993;32:66-72.

Rector AL, Rogers JE, Zanstra PE, Van Der Haring E. OpenGALEN: open source medical terminology and tools. AMIA Annu Symp Proc. 2003;:982.

Robins RH. A short history of linguistics, Longman, London 1997.

Rogers JE, Price C, Rector AL, Solomon WD, Smejko N. Validating clinical terminology structures: integration and cross-validation of Read Thesaurus and GALEN. Proc AMIA Symp. 1998;:845-9.

Scadding JG. The semantic problems of psychiatry. Psychol Med. 1990;20:243-8.

Seitinger A, Rappelsberger A, Leitich H, Binder M, Adlassnig KP. Executable medical guidelines with Arden Syntax-Applications in dermatology and obstetrics. Artif Intell Med. 2016;30321-9. https://www.ncbi.nlm.nih.gov/pubmed/27686851 (Accessed 2017/1/17)

Smith B, Kumar A, Schulze-Kremer S. Revising the UMLS Semantic Network. Medinfo. 2004;2004(CD):1700.

SNOMED. http://www.snomed.com (Accessed 2018/5/25). 2018.

Spackman KA, Campbell KE. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. Proc AMIA Symp. 1998;:740-4.

Vakiti A, Mewawalla P. Cancer, Leukemia, Myeloid, Acute (AML, Erythroid Leukemia, Myelodysplasia-Related Leukemia, BCR-ABL Chronic Leukemia). Allegheny Health Network Cancer Inst. May 14, 2018. https://www.ncbi.nlm.nih.gov/books/NBK507875/#article-25443.s2 (Accessed 2018/6/29)

Wang YY, Zhong JH, Su ZY, Huang JF, Lu SD, Xiang BD, et al. Albumin-bilirubin versus Child-Pugh score as a predictor of outcome after liver resection for hepatocellular carcinoma. Br J Surg. 2016;103:725–34. https://doi.org/10.1002/bjs.10095 (Accessed 7/12/2018).

Zhang L, Perl Y, Halper M, Geller J, Cimino JJ. An enriched unified medical language system semantic network with a multiple subsumption hierarchy. J Am Med Inform Assoc. 2004;11:195-206.