

# TriMED: banca dati terminologica multilingue

Federica Vezzani<sup>1</sup>, Giorgio Maria Di Nunzio<sup>2</sup>, Geneviève Henrot<sup>1</sup>

1 Dip. di Studi linguistici e letterari, Università di Padova, Italia - [federica.vezzani@phd.unipd.it](mailto:federica.vezzani@phd.unipd.it), [genevieve.henrot@unipd.it](mailto:genevieve.henrot@unipd.it)

2 Dip. di Ingegneria dell'Informazione, Università di Padova, Italia - [giorgiomaria.dinunzio@unipd.it](mailto:giorgiomaria.dinunzio@unipd.it)

## ABSTRACT

Tre categorie di persone si confrontano con la complessità del linguaggio medico, ciascuna con le proprie esigenze di rimedio: medici, traduttori tecnico scientifici e pazienti. Il presente lavoro propone di elaborare uno strumento che contribuisca a porre rimedio all'opacità che caratterizza la comunicazione in ambito medico tra i suoi vari attori: soddisfare la comunicazione tra pari, fornire una risorsa regolarmente aggiornata ai traduttori tecnico scientifici e facilitare la comprensione delle informazioni da parte del grande pubblico: una risorsa terminologico-fraseologica multilingue. La banca dati si compone di schede terminologiche progettate per creare un ponte fra i vari registri individuati (specialistico, semi-specialistico, non specialistico) nelle lingue considerate. Limitatamente al settore oncologico dei trattamenti per il cancro al seno, i termini da trattare sono estratti da un corpus in lingua inglese, corredati di tutte le informazioni e le proprietà linguisticamente rilevanti, e ricollegati al loro equivalente pragmatico in italiano e in francese.

## PAROLE CHIAVE

Linguaggio medico, terminologia, traduzione, metodologie e strumenti per la costruzione di risorse linguistiche.

## 1. INTRODUZIONE E STATO DELL'ARTE

Il presente intervento si inserisce nel contesto della valutazione della comunicazione in ambito sanitario e nell'analisi del linguaggio medico. Esso si caratterizza da una nomenclatura rigida e chiusa, tale da produrre un lessico opaco e carico di termini di difficile comprensione. Tra le problematiche in esso riscontrabili, le ambiguità semantiche derivano dal fenomeno della polisemia, dall'uso non corretto dei suffissi, dal mantenimento di arcaismi e dalle incongruenze etimologiche.

Numerosi studi hanno dimostrato la difficoltà a comprendere le informazioni medico-sanitarie contenute, ad esempio, nei foglietti illustrativi dei medicinali (PATEL et al., 2002), nei siti web (MCCRAY, 2005; CENTER, 2008) e più in generale nella comunicazione tra medici e pazienti (MCCRAY, 2005; JUCKS e BROMME, 2007; TRAN et al., 2009).

I pazienti sono spesso esposti ad una terminologia medica complessa: numerose ricerche pongono il proprio focus sul concetto di "comprensibilità". Lo studio di GRABAR *et al.* del 2014, ad esempio, muove un primo passo verso la semplificazione dei testi medici, in quanto propone una raccolta terminologica specifica per la valutazione di termini che rischiano di non venire compresi e che quindi richiedono ulteriori delucidazioni.

In questo quadro, ci proponiamo di costruire uno strumento terminologico multilingue che possa risolvere, per le diverse categorie di utenti individuate, i problemi correlati alla comunicazione in ambito medico.

## 2. TRE CATEGORIE DI UTENTI

In termini di divulgazione e diffusione delle conoscenze scientifiche, abbiamo identificato tre categorie di persone che, secondo diversi aspetti e a diversi livelli di comunicazione, possono risentire dei problemi collegati all'opacità del lessico medico: medici, traduttori tecnico-scientifici e pazienti.

### 2.1 MEDICI

La divulgazione su scala internazionale delle conoscenze mediche implica che la maggior parte dei testi scientifici vengano redatti in lingua inglese. In termini di diffusione di nuovi protocolli di cura e scoperte scientifiche, la lingua potrebbe pertanto ostacolare le transazioni di servizio tra colleghi medici specialisti parlanti lingue diverse, in quanto la perfetta conoscenza e padronanza della lingua straniera non è un dato scontato.

Ad un livello di comunicazione tra pari, e quindi specialistico-specialistico, ai medici serve poter valicare queste barriere linguistiche e accedere alle informazioni sulla ricerca scientifica nella propria lingua madre, in modo da importare le nuove conoscenze sul territorio nazionale, ma anche da esportare le proprie scoperte scientifiche inserendole nel circuito internazionale.

## 2.2 TRADUTTORI TECNICO-SCIENTIFICI

Nella sfera della comunicazione e della divulgazione delle informazioni, la traduzione, intesa come disciplina pratica, acquisisce un ruolo fondamentale per la corretta trasmissione di informazioni in lingue diverse. Affinché il testo tradotto risponda al principio deontologico cardine della disciplina, ovvero quello della fedeltà al testo di partenza, il traduttore tecnico-scientifico si trova a dover procedere tramite le fasi di decodificazione e transcodificazione del linguaggio medico al fine di riuscire a tradurre ciò che effettivamente comprende. Si rientra, in questo caso, in un livello di comunicazione specialistico-semi-specialistico.

Tuttavia, le esigenze del mercato della traduzione non concedono il tempo di condurre ricerche terminologiche approfondite, costringendo quindi il professionista a saltare alcuni passaggi chiave per un lavoro ottimale. Al traduttore servono risorse terminologiche regolarmente aggiornate che lo assecondino nella creazione del prodotto finale.

## 2.3 PAZIENTI

Lo sviluppo scientifico e tecnologico, che tanto influisce sulla medicina e sulla sua capacità diagnostica e terapeutica, ha progressivamente spostato il focus dell'attenzione dei medici non più sul paziente ma sulla stessa malattia, e questo ha condotto ad una crisi nel rapporto medico-paziente: il grande pubblico riscontra una difficoltà notevole nel comprendere informazioni relative alla propria salute, nonostante leggi e politiche enfatizzino la reale necessità di documentare i vari aspetti sanitari in modo più comprensibile<sup>1</sup>.

L'interazione medico-paziente implica un livello di comunicazione specialistico-non specialistico, per cui, affinché la comunicazione sia efficace, i pazienti (o più in generale il grande pubblico) necessiterebbero di comprendere il termine medico specializzato tramite l'utilizzo di una sua riformulazione in lingua "popolare" o, più in generale, tramite l'uso di un linguaggio adeguatamente tarato sul livello dell'interlocutore non-specialista.

## 3. BANCA DATI TRIMED

Le necessità precedentemente delineate per le tre categorie di utenti così identificati trovano una valida risposta in una banca terminologica multilingue. Ci siamo proposti di sviluppare un database terminologico multilingue che raccolga un insieme di schede terminologiche disegnate e compilate per ciascun termine selezionato, a ciascun livello pragmatico individuato. Disponibile gratuitamente in rete, la risorsa prende il nome di TriMED, in quanto il carattere di tripartizione è insito nello strumento: tre sono le lingue di lavoro (inglese, francese e italiano), tre sono le categorie di utenti identificati e, di conseguenza, tre sono i livelli di comunicazione oggetto di analisi dello strumento terminologico.

### 3.1 STRUTTURA

Le schede di TriMED si articoleranno su tre livelli di comunicazione:

- specializzato, fornendo sia le definizioni scientifiche per soddisfare la comunicazione tra specialisti del settore, sia il corrispondente termine tradotto per rispettare i livelli di registro linguistico;
- semi-specializzato, fornendo informazioni utili al traduttore tecnico-scientifico per la traduzione di testi settoriali;
- non-specializzato, fornendo sia definizioni divulgative per facilitare la corretta comprensione da parte del grande pubblico, sia l'equivalente del termine comunemente utilizzato nel linguaggio popolare.

---

<sup>1</sup> <https://www.kff.org/medicaid/report/ensuring-linguistic-access-in-health-care-settings-2/>

Le schede terminologiche mirano, quindi, a rendere il termine medico limpido e semanticamente accessibile anche nell'ottica della variazione diastratica.

### 3.1.1 RACCOLTA DATI

In questa fase, abbiamo compilato il corpus di partenza e i corpora paralleli e abbiamo estratto i termini tecnici.

Prima di tutto, abbiamo selezionato un insieme di articoli in lingua inglese per costituire il nostro corpus di partenza. I documenti sono stati attinti a riviste specialistiche online quali: *Breast Cancer research and treatment*<sup>2</sup>, *Archives of Breast Cancer*<sup>3</sup>, e *European Journal of cancer care*<sup>4</sup>.

Successivamente, abbiamo compilato dei corpora paralleli per le altre due lingue prese in analisi: l'italiano e il francese. Le fonti da cui abbiamo attinto i documenti in lingua italiana sono: Fondazione Umberto Veronesi (in particolare i protocolli dell'iniziativa "Pink is Good"), AIMaC - Associazione Italiana Malati di Cancro, e A.N.D.O.S - Associazione nazionale donne operate al seno. Mentre per la lingua francese: *Fondation du Cancer du Sein du Québec*, *l'Association francophone pour les soins oncologiques de support* (AFSOS) e Cairn.info.

In seguito, abbiamo estratto manualmente i termini medici per i tre corpora scelti data l'entità contenuta del corpus di partenza. Non ci siamo avvalsi del sussidio di altre risorse terminologiche disponibili sul mercato in quanto lo scopo di TriMED è quello di raccogliere e analizzare i termini presenti nei corpora di partenza selezionati.

## 3.2 APPLICAZIONE WEB

In questa sezione, applichiamo la metodologia precedentemente delineata tramite la creazione di un'applicazione Web. Per ogni termine tecnico, abbiamo provveduto a fornire le seguenti informazioni:

- equivalente in lingua popolare e divulgativa
- definizione
- analisi semica del termine
- caratteristiche formali.

Su queste informazioni abbiamo disegnato l'interfaccia TriMED che quindi si presenterà come uno strumento "a tre teste": dalla pagina iniziale, si potrà selezionare la categoria nella quale l'utente si identifica (medico, traduttore, paziente) e si potrà accedere alle relative informazioni in base alle proprie esigenze. Abbiamo implementato l'applicazione con Shiny R package (CHANG, 2015): la demo è disponibile online per mostrare il funzionamento dell'interazione fra i tre livelli di comunicazione<sup>5</sup>.

Accedendo all'interfaccia nella modalità "paziente" (Immagine 1), è possibile ricercare il termine tecnico (con suggerimento di completamento del termine) e fornire il corrispettivo in lingua popolare e la sua definizione. Nell'ottica della semplificazione del linguaggio medico, questo strumento permette quindi al paziente di "tradurre" il termine medico con una terminologia semplificata o, comunque, comunemente usata. Il paziente potrà visualizzare anche la definizione dello stesso termine. Inoltre, considerando le linee programmatiche ternarie (specialisti, semi-specialisti, non-specialisti), abbiamo pensato di fornire la possibilità di consultazione di una variante divulgativa per il termine in questione, la quale è da considerarsi diversa dal termine in lingua popolare.

Questa implementazione è ancora in fase di sviluppo e accertamento: distinguiamo tra

- termine popolare ricorrente prevalentemente in discorsi orali presso i pazienti (per esempio nelle sale d'attesa di ambulatori e ospedali), e
- termine semi-popolare (o divulgativo) usato in articoli divulgativi e che quindi estratto da corpus scritto.

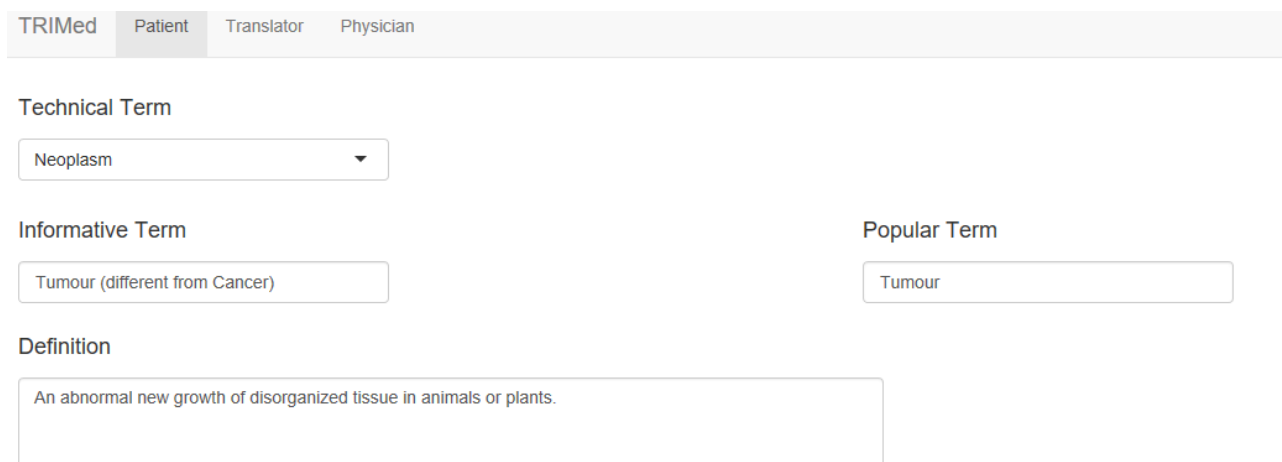
---

<sup>2</sup> <http://www.springer.com/medicine/oncology/journal/10549>

<sup>3</sup> <http://archbreastcancer.com/>

<sup>4</sup> [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1365-2354](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2354)

<sup>5</sup> <https://gmdn.shinyapps.io/TriMED/>



TRIMed Patient Translator Physician

Technical Term  
Neoplasm

Informative Term  
Tumour (different from Cancer)

Popular Term  
Tumour

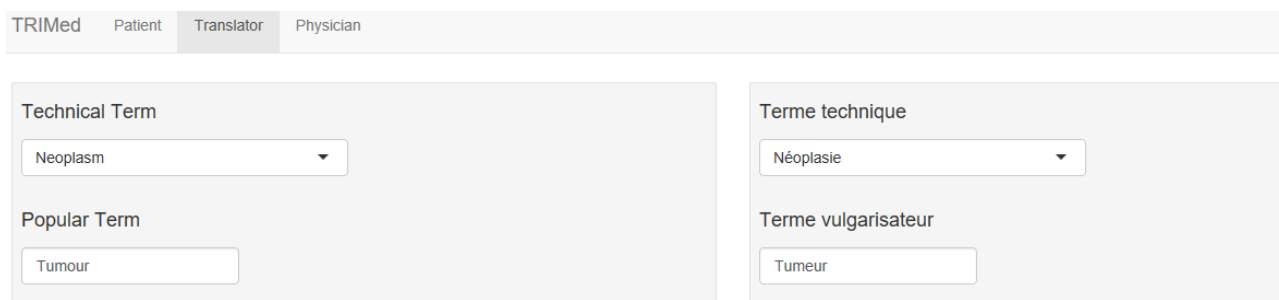
Definition  
An abnormal new growth of disorganized tissue in animals or plants.

Figura 1  
**Visualizzazione dati per l'utente di tipo paziente**

Per quanto riguarda la consultazione dell'interfaccia nella modalità “traduttore” (Figura 2 e 3), sono fornite all'utente le informazioni prettamente linguistiche e terminologiche che stanno alla base del processo traduttivo.

Dopo aver selezionato il termine, l'utente può accedere ad una schermata che si presenterà sotto forma di una tabella con duplice lettura: verticale e orizzontale.

Procedendo con un'analisi verticale, l'utente potrà consultare la traduzione del termine nella sua versione di linguaggio scientifico e popolare.



TRIMed Patient Translator Physician

Technical Term  
Neoplasm

Popular Term  
Tumour

Terme technique  
Néoplasie

Terme vulgarisateur  
Tumeur

Figura 2  
**Visualizzazione dati per l'utente di tipo traduttore**

In seguito, verrà fornita la definizione del termine e la relativa analisi semica: tale operazione ha lo scopo di scomporre il significato dell'unità lessematica o morfematica in unità atomiche non ulteriormente segmentabili.

Infine l'utente può accedere alle caratteristiche formali del termine necessarie per il suo inquadramento lessicale:

- genere;
- pronuncia;
- derivazione e composizione del termine.

#### Definition

An abnormal new growth of disorganized tissue in animals or plants.

#### Analysis

/abnormal//growth//cell//disorganized//tissue//human/

#### Features

Formal features: Part of speech: n.  
Pronunciation: /'ni:əu.plæzəm/ Other: From neo- + -plasm, after German Neoplasma.

#### Définition

Formation pathologique d'un tissu nouveau; résultat de cette formation.

#### Analyse

/formation//pathologie//tissu//corps//cellule//croissance//anormalité/

#### Caractéristiques

Caractéristiques formelles: Genre: s.f  
Dérivation/composition: Formé des élém. néo-\* et -plasie\*.

Figura 3  
**Visualizzazione dati per l'utente di tipo traduttore**

La trasmissione di queste informazioni aiuta il traduttore/la traduttrice a scegliere il candidato traduttore del termine preso in analisi.

Mantenendo la tabella con il termine di partenza, il traduttore potrà impostare la lingua di arrivo e consultare le informazioni del traduttore scelto che appariranno in una tabella adiacente a quella di partenza. In questo modo, l'utente può accedere alle stesse informazioni per il traduttore selezionato e potrà consultare orizzontalmente le informazioni per i due termini.

Infine, la modalità di accesso come utente "specialista" (Figura 4), si presenta come un'interfaccia che offre nell'ottica traduttologica l'opportunità al medico di consultare il termine nella propria lingua madre. Anche questo utente potrà selezionare la lingua di partenza e la lingua di arrivo in cui desidera consultare il termine stesso e la definizione ad esso correlata.

Per questa categoria di utente, è stato previsto anche un collegamento diretto con il MeSH term ad esso correlato. Cliccando sul termine in questione, si potrà quindi accedere alle varie informazioni fornite direttamente dalla National Library of Medicine<sup>6</sup>.

---

<sup>6</sup> <https://www.nlm.nih.gov/mesh/>

TRIMed Patient Translator **Physician**

**Technical Term**

Erythrocyte ▼

**Informative Term**

Red blood cell

**Terme technique**

Erythrocyte ▼

**Terme vulgarisateur**

Globule rouge

**Definition**

Erythrocytes are the most common type of blood cell and the vertebrate's principal means of delivering oxygen (O2) to the body tissues—via blood flow through the circulatory system.

MeSH Term: [Erythrocytes](#)

Figura 4  
**Visualizzazione dati per l'utente di tipo medico**

## 4. CONCLUSIONI

La nostra analisi delle problematiche comunicative in ambito medico ci ha portati a concepire una risorsa linguistica che fosse di aiuto per le categorie di utenti. Attualmente TriMED comprende un totale di 200 termini per le lingue francese e italiano. Stiamo lavorando sulla versione inglese in modo da adempiere agli obiettivi multilingue precedentemente delineati.

Il nostro lavoro, benché agli inizi, copre la terminologia medica nel settore oncologico e cerca di proporre una risorsa che non comprenda esclusivamente la lingua inglese ma anche altre lingue di lavoro che necessitano di terminologia documentata nel settore. Il nostro obiettivo è quindi quello di fornire uno strumento terminologico che comprenda 2500-3000 termini per questo settore specifico. Attualmente l'estrazione terminologica è stata compiuta manualmente ma prevediamo di semi-automatizzare questo processo così come l'allineamento di termini in traduzione: questo ci condurrà anche a considerazioni critiche sull'affidabilità dell'estrazione automatica in termini di granularità e pertinenza.

I dati, che attualmente sono salvati in formato tabellare, verranno modellati secondo i principi di progettazione di una base di dati relazionale. Questo tipo di organizzazione permetterà l'esportazione dei dati in vari formati. Uno degli obiettivi di questa applicazione web è quello di permettere ai traduttori tecnico-scientifici di esportare i dati in formato TermBase Exchange (TBX) che è lo standard internazionale per la rappresentazione in XML di dati terminologici<sup>7</sup>, in modo da seguire i principi del FAIR (Findable Accessible Interoperable Re-usable) Open Data della European Open Science Cloud<sup>8</sup>.

La struttura e l'idea che stanno alla base di TriMED lasciano spazio ad implementazioni future. Ad esempio, per quanto riguarda l'arricchimento della modalità "medico", ci proponiamo di permettere all'utente la visualizzazione di articoli provenienti dalle riviste scientifiche correlati a quel preciso termine tecnico. Per i termini in lingua inglese potrebbe essere utile un collegamento diretto tra i Mesh terms che rimandi a sua volta agli articoli di PubMed correlati al termine stesso. Mentre per i termini in lingua italiana, il collegamento potrebbe essere fatto con la banca dati bilingue dei Mesh a cura dell'Istituto superiore di sanità di Roma e successivo collegamento agli articoli di PubMed.

Allo stesso modo, per la terminologia francese si potrebbe procedere con il collegamento diretto al sito dell'Inserm<sup>9</sup> che in cooperazione con l'Inist-CNRS contribuisce all'aggiornamento della versione francese dei Mesh terms dal 2004.

<sup>7</sup> [http://www.ttt.org/oscarstandards/tbx/tbx\\_oscar.pdf](http://www.ttt.org/oscarstandards/tbx/tbx_oscar.pdf)

<sup>8</sup> <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

<sup>9</sup> <http://mesh.inserm.fr/mesh/>

Inoltre, considerata la reale utilità di una risorsa terminologica di questo tipo pensata nell'ottica della valutazione multilingue del linguaggio medico, ci proponiamo di ampliare le lingue di lavoro includendo anche lingue extra-europee e di estendere il settore d'interesse all'intera sfera dell'oncologia e quindi non limitandoci allo studio delle cure per il cancro al seno.

## 5. BIBLIOGRAFIA

- [1] BOUAMOR, D., LLANOS, L. C., LIGOZAT, A.-L., ROSSET, S., and ZWEIGENBAUM, P. (2016). Transferbased learning-to-rank assessment of medical term technicality. In Proceedings of the LREC 2016, Portorož, Slovenia, May. ELRA.
- [2] CENTER, O. P. (2008). Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. technical report, agency for healthcare research and quality. Oregon Evidence-based Practice Center.
- [3] CHANG, W., (2015). Shiny: Web Application Framework for R. R package version 0.11.
- [4] ELHADAD, N. and SUTARIA, K. (2007). Mining a lexicon of technical terms and lay equivalents. In Proc BioNLP Workshop, pages 49-56. ACL.
- [5] FRANÇOIS, T. and FAIRON, C. (2013). Les apports du TAL à la lisibilité du français langue étrangère. TAL, 54(1) : 171-202.
- [6] GRABAR, N., VAN ZYL, I., DE LA HARPE, R., and HAMON, T. (2014). The comprehension of medical words. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5, BIOSTEC 2014, pages 334–342, Portugal. SCITEPRESS - Science and Technology Publications, Lda.
- [7] JUCCS, R. and BROMME, R. (2007). Choice of words in doctor-patient communication: an analysis of health-related internet sites. Health Commun, 21(3):267-77.
- [8] KESELMAN, A., TSE, T., CROWELL, J., BROWNE, A., NGO, L., and ZENG, Q. (2007). Assessing consumer health vocabulary familiarity: an exploratory study. J Med Internet Res, 9(1): e5.
- [9] LEROY, G., HELMREICH, S., COWIE, J., MILLER, T., and ZHENG, W. (2008). Evaluating online health information: Beyond readability formulas. In AMIA Annual Symposium Proceedings.
- [10] MCCRAY, A. (2005). Promoting health literacy. J of Am Med Infor Ass, 12:152-163. NEVEOL, A., GROSJEAN, J., DARMONI, S., and ZWEIGENBAUM, P. (2014). Language resources for French in the biomedical domain. In Proceedings of the LREC 2014, Reykjavik, Iceland, May. ELRA.
- [11] PATEL, V., BRANCH, T., and AROCHA, J. (2002). Errors in interpreting quantities as procedures: The case of pharmaceutical labels. International journal of medical informatics, 65(3):193-211.
- [12] SERIANNI, L. (2005). Un treno di sintomi. i medici e le parole: percorsi linguistici nel passato e nel presente. Milano, Garzanti Libri.
- [13] TRAN, T., CHEKROUD, H., THIERY, P., and JULIENNE, A. (2009). Internet et soins : un tiers invisible dans la relation médecine/patient. Health Commun, 21(3): 26777.
- [14] VITALI E.D., (1967), Considerazioni sul problema del linguaggio in medicina, Protagora.
- [15] ZENG, Q., KIM, E., and CROWELL, J. and TSE, T. (2005). A text corpora-based estimation of the familiarity of health terminology. In ISBMDA, pages 184-192, Aveiro, Portugal.