CONCEPT FOR SEMANTIC ERROR ANALYSIS IN A MOBILE APPLICATION FOR SPEECH AND LANGUAGE THERAPY SUPPORT

Eugenia Rykova[1,2], Mathias Walther[1]

[1]University of Applied Sciences TH Wildau, [2]University of Eastern Finland
eugeniia.rykova@th-wildau.de

Abstract: Word finding difficulties are present in the vast majority of aphasia cases – an acquired language disorder due to focal brain injury. aphaDIGITAL project aims at developing a speech and language therapy support app for German speakers with aphasia that will provide detailed feedback on different types of errors. The current paper addresses the question of semantic errors in naming and verbal fluency tasks, building an exemplary pipeline for error analysis and providing the app user with the corresponding feedback. The pipeline is based on with GermaNet – semantic network for the German language. A way to assume words from orthographically deviating transcriptions is proposed. Limitations and further improvements are discussed.

1 Background

Aphasia, an acquired language disorder due to focal brain injury, affects all language modalities: reading and listening (comprehension), and speaking and writing (production). Speech and language therapy (SLT) improves functional communication of those who suffer from aphasia, with certain benefits brought by high intensity and duration of the therapy [1-2]. In reality, not everyone has enough access to extensive or even sufficient SLT because of geographical remoteness, lack of specialists or other reasons. Nevertheless, in-person therapy can be efficiently supplemented with digital therapy solutions used independently [3-5]. In particular, mobile applications to support SLT are becoming popular [6-7].

There is a number of typical clinical pictures of the disorder, but some linguistic symptoms can be considered as the most noticeable and universal across people with aphasia (PWA). Anomia, or word-finding problems, is present in the vast majority of aphasia cases [8]. This deficit is treated with naming-oriented semantic exercises, which can be automated with the help of automatic speech recognition (ASR). To the best of the authors' knowledge, [9-11] are currently the only three solutions that include such automation [cf. 6]. In other words, they provide feedback based on an automatically recognised oral response of the user in a nomination task. If the recognised text does not match the target, the answer of the user is accounted as incorrect without further analysis of the error nature.

In aphaDIGITAL project [12], speech and text processing algorithms are to be combined into an SLT support application for German speakers with aphasia. The app will provide detailed feedback on phonetic/phonemic, semantic and grammatical errors. The aim of the current paper is to address the question of semantic errors in picture-naming (or naming from description – both referred to as "naming task" hereinafter) and verbal fluency tasks, and build an exemplary pipeline for providing the app user with the corresponding feedback.

2 Method

2.1 Step 1: Automatic speech recognition

First, the speech input is recognized with the help of an ASR solution and compared against the target word. For this purpose, in [13] four open-source models [14-17] were selected from more than 50 ASR solutions. When compared to the target, the recognized text must pass a certain character error rate (CER) threshold. CER is calculated with the help of Python library JiWER that uses minimum-edit Levenshtein distance: when a prediction is longer than the ground truth, CER value is greater than one [18]. The threshold helps to deal with outputs deviating from standard orthography, resulting from the speech difficulties of the app user, dialect pronunciation usage, or flaws of an ASR system. Its value was provisionally set to 0.54 because it was the lowest mean CER obtained when testing the aforementioned models with aphasic speech, which allowed 72% accuracy in word verification task [13]. This value is subject to further research.

If the CER value of answer vs target comparison is lower than the threshold, the answer is considered correct and positive semantic feedback is provided to the user. If necessary, phonetic/phonematic errors are accounted, and additional remarks and correction support on pronunciation are given. When speech input does not match the target, the recognized text is passed on to semantic analysis pipeline. Currently, such pipeline is built upon GermaNet – semantic network for the German language [19].

2.2 Step 2: Semantic error analysis

Four noun word sets from the semantic category ‘food’ (in particular, subcategories ‘dairy products’, ‘fruit’, ‘vegetables’, and ‘bakery products’) were used for the experiments. Each set was composed based on the framework of a naming task, consisting of the target word in standard orthographic (St) and deviating (Dev) forms, and different types of semantically incorrect answers, also in standard and deviating forms. Most of the deviating forms are outputs of ASR models run on a small dataset obtained from four PWA during an avatar evaluation experiment: audio-samples were segmented out of the video when the participants incidentally pronounced the answers they chose on the screen [20]; the rest were artificially constructed based on this material. The sets with detailed description can be found in Table 1. For a verbal fluency task, the same sets can be used as united under the semantic (sub)category. The underlying processes of semantic analysis resemble those for a naming task.

Table 1 – Experiment word sets

Type and orthography		1. Cheese	2. Apple	3. Mushrooms	4. Biscuits
Target word	St	**Käse**	**Apfel**	**Pilze**	**Plätzchen**
Target word	Dev	Kese	Opfel	Filte	Häbchen
A word from the same semantic subcategory	St	Joghurt ‘yoghurt’	Kirsche ‘cherry’	Paprika ‘bell pepper’	Kuchen ‘cake’
A word from the same semantic subcategory	Dev	Mölch [Milch]	Oransch, Oransche [Orange]	Zwinen [Zwiebel]	Prot [Brot]
Hyponym, or a specific instance	St	Edamer ‘Edam’	Elstar ‘Elstar’	Champignon ‘champignon’	Löffelbiskuit ‘sponge finger’
Hypernym, or semantic subcategory	St	**Milchprodukt** ‘dairy product’	**Obst** ‘fruit’	**Gemüse** ‘vegetable’	**Backware** ‘bakery product’
A word from the same semantic category	St	Fleisch ‘meat’	Kartoffel ‘potato’	Keks ‘biscuit’	Schokolade ‘chocolate’
A word from the same semantic category	Dev	Preis [Reis]	Babrika [Paprika]	Rot [Brot]	Jese [Käse]
A word from a different semantic category	St	Mond ‘Moon’	Ball ‘ball’	Regenschirm ‘umbrella’	Spielzeug ‘toy’

Based on GermaNet [19], the following semantic relationships between the (incorrect) answer and the target word were to be analysed: belonging to the same semantic (sub)category and hyponymy/hypernymy. GermaNet was accessed with the academic research license

GermaNet Python API [21]. For deviating forms, close orthographic matches were searched in a list of 239,650 German words [22] with the help of Python library difflib [23]. The default threshold (0.6) was used for this search, while different maximum numbers of returned matches (n) were tested: 20, 30, 50, 100, and 500. To assume, which word could be meant by the speaker, the close matches were subject to path-based similarity analysis via GermaNet, and the semantically closest match was taken for further considerations.

3 Results

3.1 Pipeline and feedback

The following procedure is elaborated for semantic error analysis within the GermaNet framework. First, certain actions are to be performed to define the target word in the system. Some information is given by the exercise creator, and particular semantic analysis is carried out by the semantic network. Once the target word is defined in the system, it can be used for different exercises, and extra information (e.g., grammatical) can be added to it.

I. Introducing the target word by the exercise creator.
   1. Entering the target word as a lemma (dictionary form) – important for GermaNet search.
   2. Entering the target word in its target form (e.g., plural).
   3. Manual defining of the target semantic category (e.g., “food”).
   4. Defining target semantic subcategory – can be done manually or chosen later.

II. Semantic analysis of the target word carried out by the system.
   1. Proving that the target word exists in GermaNet and belongs to the given semantic category.
   2. Defining the target synonym set (synset) – a specific GermaNet unit, needed for relationship analysis.
   3. Defining target lexical category (noun, verb or adjective).
   4. Creating a list of target hypernyms up to semantic category.
   5. If target semantic subcategory was selected, proving that it is in the hypernym list, offering a hypernym list to select the subcategory otherwise.
   6. Creating a list of target hyponyms.

In a naming task, the user receives a stimulus (picture or verbal description) and says the answer, which is recorded by the system and sent to the ASR module for recognition. An exemplary pipeline of the further processing and possible feedback can be seen in Figure 1. The following actions are performed if the CER values is above the threshold and the answer is not considered correct.

III. Actions with an incorrect answer
   1. Proving if the answer is an existing (in GermaNet) word and its lexical category matches the target one.
   2. If the answer is not as an existing word, close orthographic matches are searched in a German word list, and the assumed word is selected via path-based relatedness calculator, which automatically leaves out the words form other lexical categories.
   3. If there is a word to analyse, verifying the following conditions in respect of the target:
      i. if it is in the hyponym list;
      ii. if it is in the hypernym list;
      iii. if it belongs to the same semantic subcategory;
      iv. if it belongs to the same semantic category.

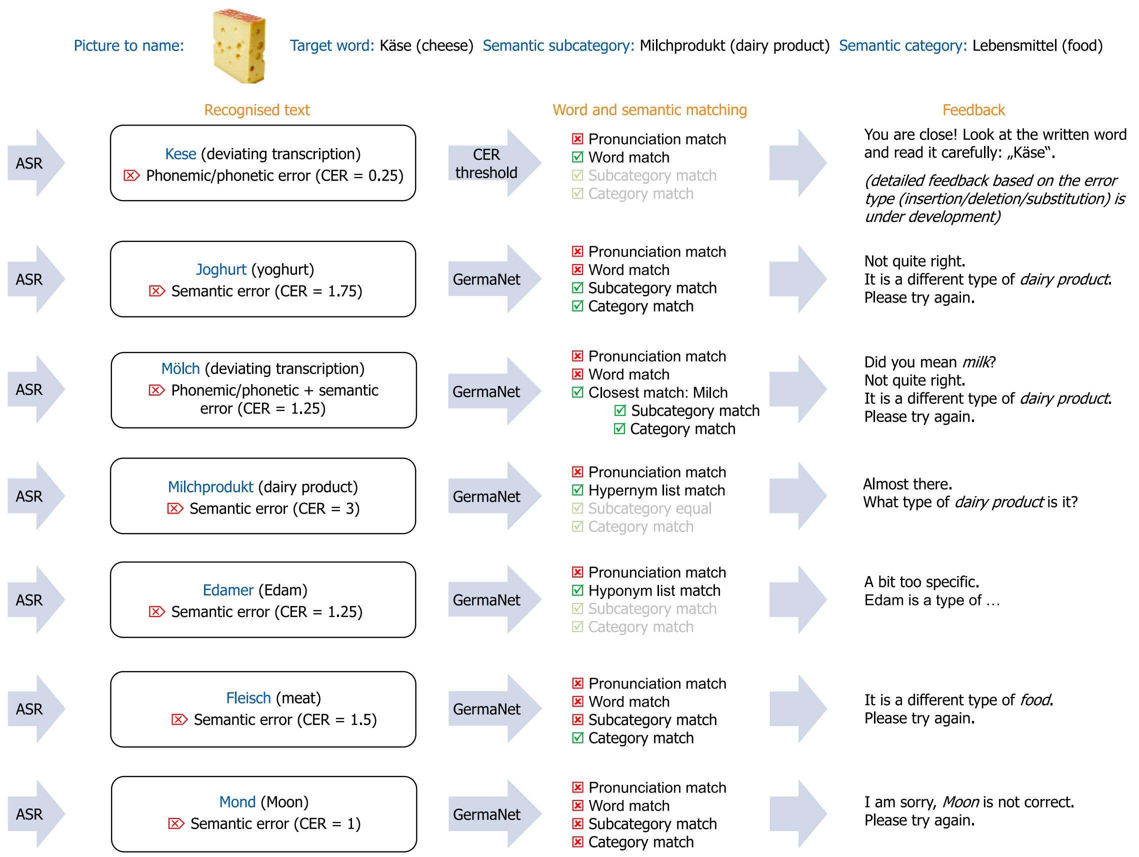

Figure 1 - Semantic error analysis and feedback exemplary pipeline

If there are no close orthographic matches, the system tells the user it could not understand what was said. If the answer or a close orthographic match is recognised as a word, but does not belong to the same lexical category as the target, the feedback on incorrectness is given. In a verbal fluency task, the list of hyponyms is created for the target (sub)category. The answers are analysed as in III.3 – in particular, if they belong to the target hyponym list. If a subcategory is used as a target, the answers can be also verified against a higher semantic category for more detailed feedback, for example "Apple is also fruit, but not citrus fruit" if the user says "apple" in a task is to name citrus fruit.

3.2 Close orthographic matches and word assuming

Table 2 displays the results of assuming a word said by the speaker, when its transcription is orthographically deviating. The words that look like an actual word (*Preis* 'price' and *Rot* 'red') are not subject to close matches search but directly analysed semantically and evaluated as wrong answers. Manual close matches search for *Preis* and *Rot* shows that the respective word *Reis* 'rice' meant by the speaker appears first when n = 500, while *Brot* 'bread' does not appear in the respective list.

From the remaining six items, two are consistently recognized as the word meant by the speaker regardless the number of close matches returned: *Mölch* > *Milch* 'milk' and *Babrika* > *Paprika* 'bell pepper'. From input *Prot*, the meant word *Brot* 'bread' is assumed while n ≤ 100, but if n = 500, the word *Printe* 'a type of gingerbread' is assumed because it is semantically closer to the target *Plätzchen* 'biscuits': simple path-based relatedness between *Brot* and *Plätzchen* is 0.91, between *Printe* and *Plätzchen* – 0.94. A similar change happens to input *Oransch*: if n = 100, the meant word *Orange* 'orange' is assumed; but if n = 500, the word *Kirsche* 'cherry' is assumed as semantically closer to the target *Apfel* 'apple': simple path-based relatedness between *Orange* and *Apfel* is 0.86, between *Kirsche* and *Apfel* – 0.89. From inputs *Zwinen* and *Oransch*, the respective meant words *Zwiebel* 'onion' and *Orange* 'orange' are assumed first when n = 500. From input *Jese*, the meant word *Käse* ‘cheese’ is not assumed with given n values, a word *Gemüse* ‘vegetable(s)’ from the same semantic category as the target is assumed when n = 500. If n = 100, the word *Wesen* ‘being’ is assumed despite the fact that *Kekse* ‘biscuits’, a synonym of the target word *Plätzchen*, is in the close matches list. In this case, it must be noted that while *Plätzchen* is both a singular and a plural form, *Kekse* is specifically a plural form, and only the singular form (lemma) *Keks* ‘biscuit’ can be found in GermaNet.

Table 2 – Automatic assumptions from orthographically deviating forms

Deviating transcription	Word meant by the speaker	Word assumed from the word list search (n – number of close matches)			
		n = 20, n =30	n = 50	n = 100	n = 500
Mölch	Milch	Milch	Milch	Milch	Milch
Oransch,	Orange	Organschaden	Organschaden	Organschaden	Orange
Oransche	Orange	Orangenschale	Flansch	Orange	Kirsche
Zwinen	Zwiebel	Zwinge	Zwinge	Zwinge	Zwiebel
Prot	Brot	Brot	Brot	Brot	Printe
Preis	Reis	*Reis* not in close matches			*Reis* in close matches
Babrika	Paprika	Paprika	Paprika	Paprika	Paprika
Rot	Brot	*Brot* not in close matches			
Jese	Käse	Jersey	Wesen	Wesen (despite Kekse)	Gemüse

4 Conclusions

In the current paper, an exemplary pipeline for semantic error analysis in a picture-naming (or naming from description) and verbal fluency tasks in German is built based on GermaNet [19]. Such analysis follows the ASR component in an SLT support mobile app aphaDIGITAL [12].

Deviations of the answer transcriptions can be overcome with the help of orthographic matches from a German word list and semantic relatedness evaluation. The current search for close matches, which results are not based on minimal editing distance but represent “the longest contiguous matching subsequence” [23], can and should be improved via introducing a CER measures to the search and subsequently adjusting the number of returned matches. First, a CER threshold can applied to the search itself. Further, a CER value and a (path-based) semantic relatedness could be weighted in the process of assuming the pronounced word to avoid such situations as *Kirsche* assumed instead of *Orange* from input *Oransche*. On the other hand, semantic relatedness can be also calculated differently, based on information content [21].

When the input is recognised as an existing word, it is automatically passed for further semantic analysis. However, an additional search for close orthographic matches seems to be reasonable to make sure it was not a phonetic/phonemic error. This consideration might be even more valid for verbal fluence task (e.g., *Preis* is recognized when *Reis* is meant in the task to name food items).

The current pipeline, or GermaNet in general, has certain limitations. First, it is only suitable for the words of the same lexical category, so that if “eat” is given as an answer to the picture of an apple, it would not be recognised as somehow related to the stimulus. Further limitations can arise from mismatch of the semantic categories in typical SLT tasks or a broader common understanding of language and GermaNet: Plätzchen ‘biscuit(s)’ could be also considered as a type of sweets (even according to the dictionary meaning), but in GermaNet this word belongs exclusively to the subcategory “bakery products”. The ways to consider the differences in grammatical forms, especially singular and plural forms for nouns, will be also explored in the following research.

List of References

[1] BHOGAL, S., R. TEASELL, AND M. SPEECHLEY: *Intensity of aphasia therapy, impact on recovery.* In *Stroke* 34(4), pp. 987–993, 2003.

[2] BRADY, M.C., H. KELLY, J. GODWIN, and P. ENDERBY: *Speech and language therapy for aphasia following stroke*. In *Cochrane Database of Systematic Reviews* 6, pp. 4–7, 2016.

[3] VAN DE SANDT-KOENDERMAN, W. M. E.: *Aphasia rehabilitation and the role of computer technology: Can we keep up with modern times?* In *International Journal of Speech-Language Pathology* 13, pp. 21-27, 2011.

[4] DES ROCHES, C. A. & S. KIRAN: *Technology-based rehabilitation to improve communication after acquired brain injury.* In *Frontiers in Neuroscience* 11:382, 2017.

[5] BRALEY, M., J. S. PIERCE, S. SAXENA, E. DE OLIVEIRA, L. TARABOANTA, V. ANANTHA, S.E. LAKHAN, and S. KIRAN: *A virtual, randomized, control trial of a digital therapeutic for speech, language, and cognitive intervention in post-stroke persons with aphasia.* In *Frontiers in Neurology*, 12:626780, 2021.

[6] GRIFFEL, J., J. LEINWEBER, B. SPELTER, and H. RODDAM: *Patient-centred design of aphasia therapy apps: a scoping review.* In *Aphasie und verwandte Gebiete | Aphasie et domaines associés*, 2, pp. 6-21, 2019.

[7] TORRE, I. G., M. ROMERO, and A. ÁLVAREZ: *Improving aphasic speech recognition by using novel semi-supervised learning methods on AphasiaBank for English and Spanish.* In *Applied Sciences* 11(19):8872, 2021.

[8] BENSON, D.: *Anomia in aphasia.* In *Aphasiology,* 2:3-4, pp. 229-235, 1988.

[9] ABAD, A., A. POMPILI, A. COSTA, I. TRANCOSO, J. FONSECA, G. LEAL, L. FARRAJOTA, and I. P. MARTINS: *Automatic word naming recognition for an on-line aphasia treatment system*. In *Computer Speech & Language* 27, pp. 1235-1248, 2013.

[10] BALLARD, K. J., N. M. ETTER, S. SHEN, P. MONROE, and C. TIEN TAN: *Feasibility of Automatic Speech Recognition for providing feedback during tablet-based treatment for apraxia of speech plus aphasia*. In *American Journal of Speech-Language Pathology* 15;28(2S), pp. 818-834, 2019.

[11] BARBERA, D. S., M. HUCKVALE, V. FLEMING, E. UPTON, H. COLEY-FISHER, C. DOOGAN, I. SHAW, W. LATHAM, A. P. LEFF, and J. T. CRINION: *NUVA: A Naming Utterance Verifier for Aphasia Treatment.* In *Computer Speech & Language* 69, 2021.

[12] APHA DIGITAL: *Entwicklung einer digitalen, dezentralen sprachtherapeutischen Versorgung [Development of digital, decentralized speech therapy solutions].* https://inno-tdg.de/projekte/aphadigital/ (last accessed 28.11.2022).

[13] RYKOVA, E. & M. WALTHER: *Evaluation of German ASR solutions for a detailed feedback in aphaDIGITAL – a mobile application for speech and language therapy support.* Manuscript submitted for publication, 2023.

[14] GROSMAN, J.: *Fine-tuned XLSR-53 large model for speech recognition in German*, 2022. https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-german (last accessed 12.09.2022).

[15] FLECK, M.: *Wav2vec2-large-xls-r-300m-german-with-lm*, 2022. https://huggingface.co/mfleck/wav2vec2-large-xls-r-300m-german-with-lm (last accessed 12.09.2022).

[16] NVIDIA: *Conformer-Transducer Large (de)*, 2022. https://huggingface.co/nvidia/stt_de_conformer_transducer_large (last accessed 12.09.2022).

[17] GUHR, O.: *Wav2vec2-large-xlsr-53-german-cv9*, 2022. https://huggingface.co/oliverguhr/wav2vec2-large-xlsr-53-german-cv9 (last accessed 12.09.2022).

[18] PYTHON SOFTWARE FOUNDATION: *JiWER: Similarity measures for automatic speech recognition evaluation*, 2022. https://pypi.org/project/jiwer/ (last accessed 15.12.2022).

[19] HAMP, B. & H. FELDWEG: *GermaNet – a lexical-semantic net for German.* In *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, 1997.

[20] ZEUNER, E., J. PIETSCHMANN, S. VOIGT-ZIMMERMANN, E. RYKOVA, and M. WALTHER: *Forschungsprojekt aphaDIGITAL: Avatar-gestützte digitale Aphasietherapie – Evaluationsstudie zur optischen und stimmlichen Modellierung des Avatars [Research project aphaDIGITAL: Avatar-supported digital aphasia therapy - evaluation study on the appearance and voice of the avatar]*. Poster presentation at: DGSS Annual Conference 2022 ‚Stimme und Geschlecht im Wandel' – Implikationen für Theorie und Praxis in der Sprechwissenschaft und Phonetik ["Voice and Gender in Transition" – Implications for Theory and Practice in Speech Science and Phonetics], Jena, Germany, 2022.

[21] TÜBINGEN UNIVERSITY SEMINAR FÜR SPRACHWISSENSCHAFT: *Python API for the German wordnet GermaNet.* https://github.com/Germanet-sfs/germanetpy (last accessed 15.11.2022).

[22] KLEUKER, D.: *Deutsche Wortliste [German word list].* https://github.com/davidak/wortliste (last accessed 08.08.2022).

[23] PYTHON SOFTWARE FOUNDATION: *difflib — Helpers for computing deltas*. https://docs.python.org/3/library/difflib.html (last accessed 08.08.2022).