Reading the Author and Speaker: Towards a Holistic and Deep Approach on Automatic Assessment of What is in One's Words

Björn W. Schuller[1,2]

[1] Department of Computing, Imperial College London,
SW7 2AZ London, UK
bjoern.schuller@imperial.ac.uk

[2] University of Passau, Chair of Complex and Intelligent Systems,
94032 Passau, Germany
http://www.schuller.one

Abstract. Computational text analysis is continuously becoming richer in ways the author of a text is 'read' in terms of the states and traits of the person behind the words such as writer's age, gender, personality, emotion or sentiment to name but a few. Similarly, in the analysis of spoken language, one finds a broadening palette of such characteristics of speakers automatically analysed in recent Computational Paralinguistics research. It seems wise to assess these characteristics in one pass to understand their interrelationship rather than going one by one in isolation. As an example, it may help to estimate one's personality knowing the age, gender, and cultural background of the person. Thus, a holistic approach is advocated that aims at automatically assessing the 'larger' picture of a person that wrote or spoke words of analysis. Here, a short motivation and inspirations 'en route' to holistic author and word-based speaker profiling are given.

Keywords: Spoken language processing · Computational paralinguistics · Speaker profiling · Text analysis · Sentiment analysis Opinion mining · Affective computing

1 What One's Words Reveal

Even when speaking or writing about others, things, or when simply telling stories that appear unrelated to ourself, our words reveal an astonishing range of attributes such as states and traits on those choosing and using them. Computational linguistics since long make use of this fact to *profile* an author of written text or speaker. Over the last decades in fact, a whole range of research challenges on the topic has emerged, including major events such as the annual author profiling task at PAN within the CLEF framework, or the sentiment analysis and other tasks in SemEval. Several further signal analysis challenges include at least the option of analysis of the spoken word, such as the annual Interspeech Computational Paralinguistics Challenge since its first edition in 2009 [42], or the annual Audio/Visual Emotion Challenge since 2011 [40], the Emotion in the Wild [15] challenge since 2013, the Multimodal Emotion Challenge since 2016 [20], and repeatedly tasks in the the ChaLearn Looking at People and MediaEval challenges, besides some more.

Here, I first want to give an overview on the diversity of author and speaker attributes computers can these days automatically derive from the spoken or written words. Methods may thereby reach from 'simple' bag-of-words approaches, e. g., learnt with support vector machines, looking at term frequencies or part-of-speech frequencies such as verb, adjective, and noun frequencies to more recent deep learnt word embeddings. They may, however, also go beyond and linguistically deeper by looking at contextual disambiguation, lexical repetition, and semantic priming effects [3], pragmatic language production abilities and deficits [17], or word-frequency mirror effects [2], and vocalisation composition [27,67].

Then, based on observations made in holistic analysis of spoken language, I want to emphasise on the need to model the author or speaker holistically, i. e., assess the different author and speaker attributes in parallel rather than in isolation. I will then, based on a dozen of taxonomies, show how the current space of attributes can literally be blown up for future holistic modelling aspirations. This will be exemplified based on the concrete tasks as were held over the years at the annual Interspeech conference in its above named challenge series focussed on paralinguistics which the author of this contribution co-organises.

2 Author and Speaker Attributes Mirrored in the Words

The automatic extraction from written or spoken text of attributes characterising and describing the person behind the words has been attempted in an astonishing richness over the last years, which I want to demonstrate next. To this end, I will use a first *taxonomy*, for a coarse categorisation: *states* and *traits* of an author or speaker. Later, further such taxonomies will be introduced aiming at 'blowing up' richness of ways to attribute an author or speaker.

2.1 Short-term States: Affect and Stances

The range of short-term states that can be extracted from one's words is impressively long; most frequently targeted ones from words include:

Affect, and valence [59], sentiment [6,43], basic emotions [45,60], continuous emotions such as arousal, dominance, and valence [39], irony [37], sarcasm [11,16], besides a sheer endless list of further states and stances, including awkwardness and assertiveness [30], disagreement [1], empathy [8], entrainment [4], flirting [29,30], friendliness [30], hostility [57], humour [23], interest [44,65], lying [24], nastiness [19], offensiveness [35], or politicalness [64].

2.2 Mid-term States: Health and Wellbeing, Mostly

When it comes to longer-term states, one can find mostly health and wellbeing related such attempted in the literature. Naturally, the boundaries between short, longer, and long-term temporal relation can be defined only loosely. For the sake of better structure and readability, here, I comprise all health and wellbeing related states, even if medical discussion may be ongoing whether some of these are rather traits as they are inherited and whether or not some or all of these are curable in theory or even practice. The list is again long, including in fore mostly Autism [36,67,68], ADHD [10], Alzheimer [2,3,18], cognitive impairment [18], communication abilities and disorders such as dietary influenced (e. g., by breast feeding or its absence) [56], depression [18,25,38,62] including shorter term concrete suicide risk [14], drug addiction [61], intoxication [65] such as from alcohol or drugs, Parkinson's [17], Rett Syndrome [27], or SAD [10].

Note that in the ongoing, I use the term *states* spanning across short to medium-term temporary characteristics of an author or speaker, such as affect or (non-chronical) health condition.

2.3 Traits: Identity, Age, Gender, and Whatnots

In contrast to the above discussed author or speaker states, *traits* are defined here by their long-term nature. Note that this does not necessarily require permanence, as for example, age (which obviously changes with the years) is subsumed under traits here.

In the first place, success for extracting traits of an individual from its words include the identity of the person encompassing its recognition and verification [9,12,63] next to the automatic identification of the age [31,55], gender [31,55], and personality [5,22,28,55]. Note that personality can also include the short-term perceived personality, such as featured in the MAPTRAITS challenge [7]. However, the author and speaker profiling literature is rich, touching upon a series of further traits [21,26,33,34].

3 Holism: Working from the Larger Picture

Holistic, by definition, lays weight on the whole and considers the interdependence of its parts. Likewise, rather than separating author and speaker attributes such as the above named states and traits and dealing with them in isolation – as is the largely dominating approach in the literature of word-based computational profiling – one should deal with the whole individual behind her words. In this section, I first want to motivate the need for a holistic approach to author and speaker profiling or characterisation by attributes; then, I show avenues from a computational view in practice, and how to further extend on holism in the future. Obviously, these are merely some inspirations – more powerful methods and models can be thought of and are yet to come.

3.1 Why to go Holistic?

The choice of our words and the deeper linguistic structure are impacted by the entirety of our self-characterising attributes such as all kinds of states and traits. When speaking or writing, we are not only doing this under the influence of, say, a specific emotion. Rather, at the same time, we may be suffering from drowsiness, stress, a cold, or even intoxication. In addition, a longer term mood or depression, and health condition can be influential factors. Obviously, the social role, stance, and the language we choose and our degree of nativeness will significantly impact on the linguistic aspects. Certainly, our age, personality, gender, social class, education, intellect, and many further factors will be further co-influencing this choice of words and structure. Yet, besides few examples as a study investigating the impact of emotion on author profiling [32], the vast majority of computational language processing literature ignores this co-dependence of attributes and focusses on one aspect at a time – such as emotion in this example – risking severe downgrades in accuracy once applied in real world automatic recognitions applications, where data comes from real-life highly blended out-of-the-lab situations.

3.2 How to go Holistic: Methods

A number of approaches exist in the maching learning body of literature to assess multiple attributes such as the ones described above synergistically either iteratively or in full parallel. Typical examples include neural networks with multiple output neurons for accordingly several targets such as classes or regression of several tasks. But a plethora of approaches also exists for other types of learning algorithms such as Support Vector Machines [54].

A major bottleneck for training of such approaches is the lack of data labelled in a rich variety of attributes of authors and speakers. As re-labelling of data may be tedious and labour intensive, and partially simply not possible, as information on the ground truth states and traits may not be accessible for a broad range of these, semi automatic and fully automatic approaches were introduced to relabel databases one by another. An example is cross-task labelling (CTL) as introduced in [69]. Transfer learning can be a further alternative to learn across tasks and conditions [13] and enrich one's database in largely automated ways.

3.3 How to go Holistic: More Taxonomies

Up to this point, in Sect. 2 we were considering a single taxonomy to group or classify author and speaker attributes of interest, namely states vs traits. However, one can group these in a range of further ways [41], which will introduce also new viewing angles in terms of holism. Strictly speaking, using additional taxonomies will not increase the number of attributes one may target, but can help make their organisation easier, if we consider, e. g., acted vs spontaneous as an additional second taxonomy. As an example, we could add 'acted pain' and 'spontaneous pain' as further states of interest, but if 'pain' is handled as

Table 1. A dozen of taxonomies: frequent and important options for the grouping of author and speaker attributes with a short comment. See [41] for a detailed explanation.

Taxonomy	Comment
Trait vs state	Relates to time/permanence
Acted vs spontaneous	One may further consider masking or regulation
Complex vs simple	Blended or 'pure'
Measured vs assessed	Relates to objectivity of the 'ground truth'
Categorical vs continuous	e. g., a set of emotions vs continuous emotions
Felt vs perceived	Perceived by observers, i. e., others
Intentional vs instinctual	e. g., acting is usually intentional
Consistent vs discrepant	e. g., irony being discrepant
Private vs social	Relating to the communication intention
Prototypical vs peripheral	Salient, central example or unusual or atypical
Universal vs culture-specific	e. g., affect or acting may depend on the culture
Uni-modal vs multi-modal	Here: linguistics only or including acoustics

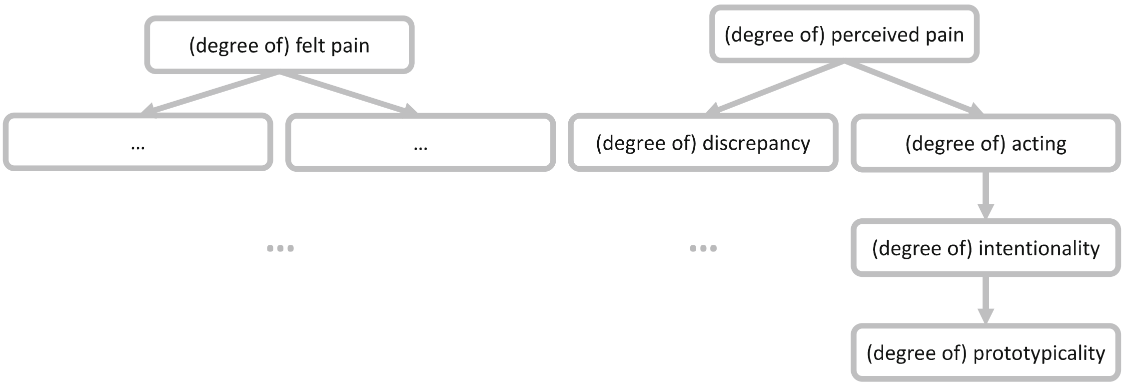

Fig. 1. An example of a semantic tree for '(degree of) pain)' inspired by the taxonomies in Table 1. A higher order single root node representing the author or speaker was left out for better visibility. Explanations are given in the text.

a regression problem in the sense of 'degree of pain', it may be preferable to have a pain attribute and a related degree of acting attribute. One may now attach further related attributes and attach a property describing the cultural connection, describing to which culture the signals of pain or its acting relate. Likewise, one can think of a semantic tree structure with a root attribute – in the example '(degree of) pain' – and subsequent related attributes that describe the type of the root aspect (such as '(degree of) pain'). In Table 1, the taxonomies as were introduced in [41] – where one can find very detailed explanations of these – are given alongside a short comment for their short 'in a nutshell'-type familiarisation.

In Fig. 1, an example is given on how one may translate these taxonomies to possible ways of forming semantic trees as indicated to reach much richer descriptions of speaker attributes than current approaches in the literature target. The root nodes '(degree of) felt pain' and '(degree of) perceived pain' show, how a taxonomy was used to build two different attributes which, in a machine learning approach, could be learnt as multiple targets given their likely high correlation. One would reflect the actual pain the individual experiences (left tree), the other the pain others would perceive within the individual (right tree). Note that, one could add a new root 'author' or 'speaker' above to form a single tree, which was left out here for better visibility. On the first layer as shown in the figure, in a truly holisic approach, one would find all sorts of further attributes, e. g., grouped by states and traits, characterising the subject. Then, on the second layer in the figure, only the right hand tree is filled with exemplary further taxonomies for better visibility. Here, we find exemplary attributes that describe this perception of the author's or speaker's '(degree of) pain' by others: On the left-hand side, the '(degree of) discrepancy', on the right hand side, the '(degree of) intentionality'; up to this point, following the right hand side, we would have a description on how the portrayal of pain by a subject is perceived by others – as acted or real, and if they perceive this degree of acting as instinctual or intentional. To go one layer deeper, we could next add how prototypical this example would be by the '(degree of) prototypicality. This would give us, for example, a sample of text or speech that could be labelled as a 'typical example of intentionally acted pain in observers' eyes. Obviously, despite the figure showing dots at several places, it seems hardly reasonable to either go for full depth nor for full width, as one would need labelling for any of these aspects. However, it is interesting to see how one can likewise find a new way of describing author and speaker attributes as compared to today's dominating approach of targeting single aspects in isolation.

3.4 How to go Holistic: Adding Acoustics

Speaking of spoken language analysis, it seems wise to also consider analysis of the acoustic properties. This may also comprise adding acoustic confidences in the linguistic analysis coming from an automatic speech recognition engine. In fact, many automatic speaker attribute assessment tasks in spoken language analysis show good synergy when exploiting both acoustic and linguistic information. Examples include most notably emotion [42]. Major challenges for the exploitation of both information types thereby include the often smaller size of corpora for acoustic model training as compared to such desired and typically used in linguistic model training; also, not all corpora as used for acoustic model training can be used for linguistic analysis, as these are partially based on prompted text or even just vowel-consonant combinations. Further, fusion of the information streams can be less straight forward given the different time levels these operate on: for linguistic analyses, one mostly considers larger amounts of text than usual chunk sizes of one or a few words as are used for acoustic analysis would be – however, early fusion on the feature level has repeatedly been shown to be feasible [41]. The optimal type of fusion itself is not at all decided upon, either [41].

Table 2. Overview on the tasks of the Interspeech Computational Paralinguistics and according pre-decessor challenges centred on acoustic analysis of paralinguistic effects, while often allowing also for linguistic assessment. For simplification, a binary classification is made per taxonomy as commented upon in Table 1 and the data as was used in the according (sub-)challenge; note, however, that instead, one could also introduce a continuous dimension per taxonomy as shown in Fig. 1. In fact, this coarse discretisation of two classes renders some decisions rather ambiguous, and other classifications could partially be assigned. Also note that the assignment here does not hold for the author or speaker attribute in general (i. e., the 'phenomenon'), but is focussed on the specific task/data of the according (sub-)challenge as was held in the according year (cf. last column). '+' denotes cases where both options exist in the data. In the case of c/n (categorical or continuous), a '+' is given if the data includes continuous annotation, despite a classification task by discretisation was used in the challenge. In the case of u/m (unimodal vs multimodal), here, 'u' is given if the attributes can be inferred only by acoustics, 'm' is given, if use of linguistic analysis is reasonable, and '+' is given if u/m hold for different parts of the database. The language(s) of the content are additionally given by country code (ISO 3166 ALPHA 2). En: English (w/o specific region). Ps: Pseudo-language.

Taxonomy	trait/state	acted/spontaneous	complex/simple	measured/assessed	categorical/continuous	felt/perceived	intentional/instinctual	consistent/discrepant	private/social	prototypical/peripheral	universal/culture-specific	uni-modal/multi-modal	
Abbreviation	t/s	a/s	c/s	m/a	c/n	f/p	i/n	c/d	p/s	p/n	u/c	u/m	Lang. Year
Addressee	s	s	s	a	c	p	+	+	s	+	c	m	US 2017
Age	t	s	s	m	+	f	n	c	+	+	u	m	DE 2010
Autism	t	s	s	m	c	f	n	c	p	+	u	u	FR 2013
Cognitive Load	s	s	s	m	+	f	n	c	p	+	c	u	AU 2014
Cold	s	s	s	a	+	f	n	c	p	+	u	+	DE 2017
Conflict	s	s	s	a	+	p	+	+	+	+	c	m	FR 2013
Deception	s	s	s	m	c	f	+	+	p	+	c	m	US 2016
Eating	s	s	s	m	+	f	+	c	p	+	u	+	DE 2015
Emotion (acted)	s	a	s	m	c	p	i	+	s	p	c	u	Ps 2013
Emotion (spontaneous)	s	s	c	a	c	p	+	+	+	+	c	m	DE 2009
Gender	t	s	s	m	c	f	n	c	+	+	c	+	DE 2010
Pathology	t	s	s	a	+	p	n	c	p	+	u	u	NL 2012
Interest	s	s	s	a	n	p	+	+	+	+	c	m	En 2010
Intoxication	s	s	s	m	+	f	n	c	p	+	c	+	DE 2011
Likability	t	s	s	a	+	p	+	+	+	+	c	+	DE 2012
Native Language	t	s	s	m	c	f	n	c	+	+	u	+	En 2016
Nativeness	s	s	s	a	n	p	+	+	+	+	c	+	En 2015
Parkinson's	s	s	s	a	n	p	n	c	p	+	u	+	ES 2015
Personality	t	s	s	a	+	p	+	+	+	+	c	m	FR 2012
Physical Load	s	s	s	m	+	f	n	c	p	+	u	u	DE/En 2014
Sincerity	s	a	s	a	n	p	+	+	s	p	c	u	En 2016
Sleepiness	s	s	s	a	+	+	n	c	p	+	c	+	DE 2011
Snoring	s	s	s	a	c	p	n	c	p	+	u	u	– 2017
Social Signals	s	s	s	m	c	+	+	+	+	+	c	m	UK 2013

However, not all tasks have been attempted exploiting each of these two information types (acoustics and linguistics) in isolation, i. e., exclusively by acoustic, *or* linguistic information. This comes, as some speaker attributes are rather unsuited to be assessed by each of the two types of information at a similarly high precision. As an example, consider the case of speaker height recognition [66]: while height correlates in a certain age range with age, which clearly has an influence on linguistics, it will mostly manifest in acoustic features for adult speakers.

Other tasks have been attempted by both of the two types – acoustics and linguistics – also mutually exclusively such as race [58]. However, different potential can partially be observed, such as in the case of emotional speech modelling, arousal being better assessed by acoustic cues, and valence better be assessed by linguistic cues [41]. This leads to the question, when to preferably use which or both of these two types of information. As a basic assumption, one may assume that aspects which reflect physical differences reflect more or only in acoustics, whereas cognitive aspects reflect more in linguistic cues. However, in a real-world, correlation and co-effects may benefit each of the two, such as when trying to recognise if a speaker is a smoker or not. While a smoker's voice clearly differs acoustically from the one of a non-smoker, one would not assume that a smoker should linguistically differ from a non-smoker. However, depending on the test sample, or even distribution in the broader population, other effects may come into play that *correlate* with being a smoker or not, *and* impact on linguistics. For example, smokers in a test sample of a database might represent different age classes depending on when smoking was more or less popular, or represent different social classes, gender, or alike, which as a phenomenon all do impact on linguistics. Note that in the ongoing, the above introduced taxonomy 'unimodal vs multimodal' (cf. Table 1) which alludes to whether an attribute relates to one or multiple modalities will, looking at spoken language, reflect whether acoustic and linguistic phenomena can be handled only by one or by both of these types of information.

Let us now look at a concrete example of speaker attributes for further exemplification of the above principles: Based on the Interspeech Computational Paralinguistics Challenge (ComParE) series [46,48–50,53] (including its predecessors on emotion [42], paralinguistics [47], speaker state [52], and speaker trait [51][1], the tasks of speaker attribution as given in Table 2 were assessed by both cues (acoustics and linguistics) during the according challenge or seem to be suited to be attempted by linguistics in general.

The table provides a coarse breakdown of the specific tasks and data as were held in the challenge's sub-challenges. Likewise, note that it does not categorise the author or speaker attributes per se, but indeed the specific conditions of the data as were held. For example, in the case of acted or spontaneous, eating is marked as spontaneous. One could also act speech under eating, but during the recording of this data, actual eating was given – thus, it is marked as spontaneous in the table. A certain ambiguity here lies in the fact that, the eating

[1] More information can be accessed from http://www.compare.openaudio.eu.

took place in a lab with normalised intake conditions, thus, not feeling 'spontaneous', yet, clearly not being acted. Similarly, one could act a native language or sleeping, to name but two – however, in the according sub-challenge tasks, this was not the case, thus leading to the spontaneous label in the table. This is just an exemplification of the ambiguity coming with making only binary decisions. However, the main point of the table is to demonstrate the huge richness one can find in applying the dozen taxonomies shown in both tables: While the challenge dealt for example with speech under eating in 2015 as is shown in the Table 2, it specifically only dealt with non-acted eating. Accordingly, one could now consider 'speech under acted eating' as a novel attribution. Following this principle, one could now massively extend the holistic space of possible author and speaker attributes by looking at all missing combinations in the table, and – of course – by adding many more attributes in the first place.

4 Summary and Conclusion

Novel concepts of rich and structured author or speaker attributes were introduced. Repeatedly targeted dominating examples of (flat) attributes were further given that show the current state of automatic assessment from written or spoken language which is dominated by isolated handling of these. Based on these, first, a three-fold division was made into short-term and medium-term states and long-term traits to group these. Then, further taxonomies for grouping and conceptualisation were shown. Two main points were then raised:

First, author and speaker attributes should be assessed holistically in their commonality, as they largely and often co-influence each other. This is also important, as otherwise, it remains unclear what is actually recognised once going for automatic processing 'in the wild' in real-world applications. As an example, both speaking under a cold or while eating may have an impact on linguistics, as with a soar throat or while chewing one may speak in short phrases and an emphasis on specific sounds to avoid speaking too much and too strenuously. Likely, this will also influence the composition of adjectives, nouns, verbs, and other part-of-speech classes one chooses. However, a study looking only at one of these two phenomena versus 'normal speech' may be overly optimistic in its results and performance assumptions as in a real world application, both cases will happen over time, making confusions or false positives very likely. To tackle this problem, examples of ways of co-learning and multi-target classification and regression were shortly given. These included methods based on semi-supervised and transfer learning to re-label data in other databases' labels, as richly annotated let alone 'holistically' annotated data is scarce.

Second, fusion of acoustic and linguistic information in the sense of another 'holistic view' on spoken language analysis is broadly considered as synergistic, yet, for many states and traits it has hardly been attempted. A number of obstacles were named including differences in corpora usually existent in these sub-disciplines, different timing levels of analysis, and the ever-ongoing discussion of the optimal way to fuse these information sources.

To exemplify the current situation and indicate future avenues of broadening on holism, a snapshot image was given by the Interspeech Computational Paralinguistics Challenge series which is focussed on acoustic analysis, but also allows for linguistic analysis. There, one could first see the potential of not yet attempted spoken language analysis tasks by either only linguistic approaches or a combination of these with acoustic processing. Second, as these tasks were classified by a dozen of taxonomies, one could see how the kinds of attributes could be massively en-richened such as in an also introduced tree structure.

Overall, this leaves an interesting field of research efforts to mine for the future, in which intelligent systems will be able to profile authors and speakers in a broad range of attributes grouped in many ways such as by states and traits in full parallel seeing and hearing the 'larger picture' of the person behind the written and spoken words exploiting also the acoustic channel if available. One may wonder how oncoming machines equipped with such rich emotional and social intelligence may interact with us, retrieve information from spoken and written language in so far unseen richness and accuracy, but also monitor us for wellbeing, and – ultimately – change future society technology. May it be used only for society's best.

Acknowledgement. The author would like to thank his colleague Anton Batliner for discussion regarding the taxonomies as discussed herein. He further acknowledges funding from the European Research Council within the European Union's 7th Framework Programme under grant agreement no. 338164 (Starting Grant Intelligent systems' Holistic Evolving Analysis of Real-life Universal speaker characteristics (iHEARu)), and the European Union's Horizon 2020 Framework Programme under grant agreement no. 645378 (Research Innovation Action Artificial Retrieval of Information Assistants – Virtual Agents with Linguistic Understanding, Social skills, and Personalised Aspects (ARIA-VALUSPA)). The responsobility lies with the author.

References

1. Abbott, R., Walker, M., Anand, P., Fox Tree, J.E., Bowmani, R., King, J.: How can you say such things? recognizing disagreement in informal political argument. In: Proceedings Workshop on Languages in Social Media, pp. 2–11. ACL, Portland, Oregon (2011)
2. Balota, D.A., Burgess, G.C., Cortese, M.J., Adams, D.R.: The word-frequency mirror effect in young, old, and early-stage alzheimer's disease: evidence for two processes in episodic recognition performance. J. Mem. Lang. **46**(1), 199–226 (2002)
3. Balota, D.A., Duchek, J.M.: Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the alzheimer type. Brain Lang. **40**(2), 181–201 (1991)
4. Beňuš, Š.: Social aspects of entrainment in spoken interaction. Cogn. Comput. **6**(4), 802–813 (2014)
5. Biel, J.I., Tsiminaki, V., Dines, J., Gatica-Perez, D.: Hi youtube!: personality impressions and verbal content in social video. In: Proceedings ICMI, pp. 119–126. ACM, Sydney, Australia (2013)

6. Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New avenues in opinion mining and sentiment analysis. IEEE Intell. Syst. Mag. **28**(2), 15–21 (2013)
7. Celiktutan, O., Eyben, F., Sariyanidi, E., Gunes, H., Schuller, B.: MAPTRAITS 2014: the first audio/visual mapping personality traits challenge. In: Proceedings Personality Mapping Challenge & Workshop (MAPTRAITS), Satellite of ICMI, pp. 3–9. ACM, Istanbul, Turkey (2014)
8. Chakravarthula, S.N., Xiao, B., Imel, Z.E., Atkins, D.C., Georgiou, P.G.: Assessing empathy using static and dynamic behavior models based on therapist's language in addiction counseling. In: Proceedings Interspeech, pp. 668–672. ISCA, Dresden, Germany (2015)
9. Chaski, C.E.: Empirical evaluations of language-based author identification techniques. Forensic Linguist. **8**, 1–65 (2001)
10. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K.: From adhd to sad: analyzing the language of mental health on twitter through self-reported diagnoses. In: Proceedings CLPsych at NAACL HLT, p. 1. ACL, Denver, Colorado (2015)
11. Davidov, D., Tsur, O., Rappoport, A.: Semi-supervised recognition of sarcastic sentences in twitter and amazon. In: Proceedings 14th Conference on Computational Natural Language Learning, pp. 107–116. ACL, Uppsala, Sweden (2010)
12. De Vel, O., Anderson, A., Corney, M., Mohay, G.: Mining e-mail content for author identification forensics. ACM SIGMOD Rec. **30**(4), 55–64 (2001)
13. Deng, J., Xu, X., Zhang, Z., Frühholz, S., Schuller, B.: Universum autoencoder-based domain adaptation for speech emotion recognition. IEEE Signal Process. Lett. **24**, 5 (2017)
14. Desmet, B., Pauwels, K., Hoste, V.: Online suicide risk detection using automatic text classification. In: Proceedings International Summit on Suicide Research. IASR/AFSP, New York City, 1 p. (2015)
15. Dhall, A., Goecke, R., Joshi, J., Wagner, M., Gedeon, T.: Emotion recognition in the wild challenge 2013. In: Proceedings ICMI, pp. 509–516. ACM, Sydney, Australia (2013)
16. González-Ibánez, R., Muresan, S., Wacholder, N.: Identifying sarcasm in twitter: a closer look. In: Proceedings ACL, vol. 2, pp. 581–586. ACL, Portland, Oregon (2011)
17. Holtgraves, T., Fogle, K., Marsh, L.: Pragmatic language production deficits in Parkinson's disease. In: Advances in Parkinson's Disease, vol. 2, pp. 31–36 (2013)
18. Jarrold, W.L., Peintner, B., Yeh, E., Krasnow, R., Javitz, H.S., Swan, G.E.: Language analytics for assessing brain health: cognitive impairment, depression and pre-symptomatic alzheimer's disease. In: Yao, Y., Sun, R., Poggio, T., Liu, J., Zhong, N., Huang, J. (eds.) BI 2010. LNCS (LNAI), vol. 6334, pp. 299–307. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15314-3_28
19. Justo, R., Corcoran, T., Lukin, S.M., Walker, M., Torres, M.I.: Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web. Knowl. Based Syst. **69**, 124–133 (2014)
20. Li, Y., Tao, J., Schuller, B., Shan, S., Jiang, D., Jia, J.: MEC 2016: the multimodal emotion recognition challenge of CCPR 2016. In: Tan, T., Li, X., Chen, X., Zhou, J., Yang, J., Cheng, H. (eds.) CCPR 2016. CCIS, vol. 663, pp. 667–678. Springer, Singapore (2016). https://doi.org/10.1007/978-981-10-3005-5_55
21. Lin, J.: Automatic author profiling of online chat logs. Ph.D. thesis, Naval Postgraduate School, Monterey, California (2007)
22. Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using linguistic cues for the automatic recognition of personality in conversation and text. J. Artif. Intell. Res. **30**, 457–500 (2007)

23. Mihalcea, R., Strapparava, C.: Making computers laugh: investigations in automatic humor recognition. In: Proceedings Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 531–538. ACL, Vancouver, Canada (2005)
24. Mihalcea, R., Strapparava, C.: The lie detector: explorations in the automatic recognition of deceptive language. In: Proceedings ACL-IJCNLP Conference Short Papers, pp. 309–312. ACL, Singapore (2009)
25. Neuman, Y., Cohen, Y., Assaf, D., Kedma, G.: Proactive screening for depression through metaphorical and automatic text analysis. Artif. Intell. Med. **56**(1), 19–25 (2012)
26. Patra, B.G., Banerjee, S., Das, D., Saikh, T., Bandyopadhyay, S.: Automatic author profiling based on linguistic and stylistic features. Notebook for PAN at CLEF 1179 (2013)
27. Pokorny, F.B., Marschik, P.B., Einspieler, C., Schuller, B.W.: Does she speak RTT? Towards an earlier identification of RETT syndrome through intelligent prelinguistic vocalisation analysis. In: Proceedings Interspeech, pp. 1953–1957 (2016)
28. Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., Howard, N.: Common sense knowledge based personality recognition from text. In: Castro, F., Gelbukh, A., González, M. (eds.) MICAI 2013. LNCS (LNAI), vol. 8266, pp. 484–496. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-45111-9_42
29. Ranganath, R., Jurafsky, D., McFarland, D.: It's not you, it's me: detecting flirting and its misperception in speed-dates. In: Proceedings Conference on Empirical Methods in Natural Language Processing, vol. 1, pp. 334–342. ACL, Singapore (2009)
30. Ranganath, R., Jurafsky, D., McFarland, D.A.: Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates. Comput. Speech Lang. **27**(1), 89–115 (2013)
31. Rangel, F., Rosso, P.: Use of language and author profiling: identification of gender and age. In: Natural Language Processing and Cognitive Science, vol. 177 (2013)
32. Rangel, F., Rosso, P.: On the impact of emotions on author profiling. Inf. Process. Manag. **52**(1), 73–92 (2016)
33. Rangel, F., Rosso, P., Koppel, M.M., Stamatatos, E., Inches, G.: Overview of the author profiling task at pan 2013. In: Proceedings CLEF Conference on Multilingual and Multimodal Information Access Evaluation, pp. 352–365. CELCT, Amsterdam, The Netherlands (2013)
34. Rangel, F., Rosso, P., Potthast, M., Stein, B., Daelemans, W.: Overview of the 3rd author profiling task at pan 2015. In: Proceedings CLEF, 18 p. CLEF Association, Toulouse, France (2015)
35. Razavi, A.H., Inkpen, D., Uritsky, S., Matwin, S.: Offensive language detection using multi-level classification. In: Farzindar, A., Kešelj, V. (eds.) AI 2010. LNCS (LNAI), vol. 6085, pp. 16–27. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13059-5_5
36. Regneri, M., King, D.: Automatically evaluating atypical language in narratives by children with autistic spectrum disorder. In: Sharp, B., Delmonte, R. (eds.) Natural Language Processing and Cognitive Science: Proceedings 2014, pp. 173–186. Walter de Gruyter GmbH & Co KG, Berlin (2015)
37. Reyes, A., Rosso, P., Buscaldi, D.: From humor recognition to irony detection: the figurative language of social media. Data Knowl. Eng. **74**, 1–12 (2012)
38. Rude, S., Gortner, E.M., Pennebaker, J.: Language use of depressed and depression-vulnerable college students. Cogn. Emot. **18**(8), 1121–1133 (2004)

39. Schuller, B.: Recognizing affect from linguistic information in 3D continuous space. IEEE Trans. Affect. Comput. **2**(4), 192–205 (2012)
40. Schuller, B.: The computational paralinguistics challenge. IEEE Signal Process. Mag. **29**(4), 97–101 (2012)
41. Schuller, B., Batliner, A.: Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. Wiley, U.S. (2013)
42. Schuller, B., Batliner, A., Steidl, S., Seppi, D.: Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge. Speech Commun. **53**(9/10), 1062–1087 (2011)
43. Schuller, B., Mousa, A.E.D., Vasileios, V.: Sentiment analysis and opinion mining: on optimal parameters and performances. WIREs Data Min. Knowl. Discov. **5**, 255–263 (2015)
44. Schuller, B., Müller, R., Hörnler, B., Höthker, A., Konosu, H., Rigoll, G.: Audiovisual recognition of spontaneous interest within conversations. In: Proceedings ICMI, pp. 30–37. ACM, Nagoya, Japan (2007)
45. Schuller, B., Rigoll, G., Lang, M.: Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In: Proceedings ICASSP, vol. 1, pp. 577–580. IEEE, Montreal, Canada (2004)
46. Schuller, B., etal.: The INTERSPEECH 2017 computational paralinguistics challenge: addressee, cold & snoring. In: Proceedings Interspeech, 5 p. ISCA, Stockholm, Sweden (2017)
47. Schuller, B., et al.: Paralinguistics in speech and language—state-of-the-art and the challenge. Comput. Speech Lang. **27**(1), 4–39 (2013)
48. Schuller, B., et al.: The INTERSPEECH 2014 computational paralinguistics challenge: cognitive & physical load. In: Proceedings Interspeech. ISCA, Singapore (2014)
49. Schuller, B., et al.: The INTERSPEECH 2015 computational paralinguistics challenge: degree of nativeness, parkinson's & eating condition. In: Proceedings Interspeech, pp. 478–482. ISCA, Dresden, Germany (2015)
50. Schuller, B., et al.: The INTERSPEECH 2016 computational paralinguistics challenge: deception, sincerity & native language. In: Proceedings Interspeech, pp. 2001–2005. ISCA, San Francisco, California (2016)
51. Schuller, B., et al.: A survey on perceived speaker traits: personality, likability, pathology, and the first challenge. Comput. Speech Lang. **29**(1), 100–131 (2015)
52. Schuller, B., et al.: Medium-term speaker states—a review on intoxication, sleepiness and the first challenge. Comput. Speech Lang. **28**(2), 346–374 (2014)
53. Schuller, B., et al.: The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: Proceedings Interspeech, pp. 148–152. ISCA, Lyon, France (2013)
54. Schuller, B., Zhang, Y., Eyben, F., Weninger, F.: Intelligent systems' holistic evolving analysis of real-life universal speaker characteristics. In: Proceedings 5th International Workshop on Emotion Social Signals. Sentiment & Linked Open Data (ES^3LOD 2014), Satellite of LREC, pp. 14–20. ELRA, Reykjavik, Iceland (2014)
55. Schwartz, H.A., et al.: Personality, gender, and age in the language of social media: the open-vocabulary approach. PloS ONE **8**(9), e73791 (2013)
56. Smith, J.M.: Breastfeeding and language outcomes: a review of the literature. J. Commun. Disord. **57**, 29–40 (2015)
57. Spertus, E.: Smokey: automatic recognition of hostile messages. In: Proceedings AAAI/IAAI, pp. 1058–1065. AAAI Press, Providence, Rhode Island (1997)

58. Squires, G.D., Chadwick, J.: Linguistic profiling a continuing tradition of discrimination in the home insurance industry? Urban Aff. Rev. **41**(3), 400–415 (2006)
59. Strapparava, C., Mihalcea, R.: Semeval-2007 task 14: affective text. In: Proceedings 4th International Workshop on Semantic Evaluations, pp. 70–74. ACL, Prague, Czech Republic (2007)
60. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: Proceedings ACM Symposium on Applied Computing, pp. 1556–1560. ACM, Fortaleza, Brazil (2008)
61. Strapparava, C., Mihalcea, R.: A Computational Analysis of the Language of Drug Addiction, vol. 2, pp. 136–142 (2017)
62. Valstar, M., Gratch, J., Schuller, B., Ringeval, F., Cowie, R., Pantic, M.: Summary for AVEC 2016: depression, mood, and emotion recognition workshop and challenge. In: Proceedings ACM International Conference on Multimedia, pp. 1483–1484. ACM, Amsterdam, The Netherlands (2016)
63. Van Halteren, H.: Linguistic profiling for author recognition and verification. In: Proceedings ACL, p. 199. ACL, Barcelona, Spain (2004)
64. Vijayaraghavan, P., Vosoughi, S., Roy, D.: Automatic detection and categorization of election-related tweets. In: Proceedings 10th International AAAI Conference on Web and Social Media. AAAI, Cologne, Germany (2016)
65. Wang, W.Y., Biadsy, F., Rosenberg, A., Hirschberg, J.: Automatic detection of speaker state: lexical, prosodic, and phonetic approaches to level-of-interest and intoxication classification. Comput. Speech Lang. **27**(1), 168–189 (2013)
66. Weninger, F., Wöllmer, M., Schuller, B.: Automatic assessment of singer traits in popular music: gender, age, height and race. In: Proceedings ISMIR, pp. 37–42. ISMIR, Miami, FL, USA (2011)
67. Xu, D., Gilkerson, J., Richards, J., Yapanel, U., Gray, S.: Child vocalization composition as discriminant information for automatic autism detection. In: Proceedings EMBC, pp. 2518–2522. IEEE, Minneapolis (2009)
68. Xu, D., Richards, J.A., Gilkerson, J., Yapanel, U., Gray, S., Hansen, J.: Automatic childhood autism detection by vocalization decomposition with phone-like units. In: Proceedings 2nd Workshop on Child, Computer and Interaction, Satellite of ICMI-MLMI, p. 5. ACM, Cambridge, MA, USA (2009)
69. Zhang, Y., Zhou, Y., Shen, J., Schuller, B.: Semi-autonomous data enrichment based on cross-task labelling of missing targets for holistic speech analysis. In: Proceedings ICASSP, pp. 6090–6094. IEEE, Shanghai, P.R. China (2016)