Clinical Linguistics & Phonetics, August 2012; 26(8): 661–680

ISSN: 0269-9206 print / ISSN 1464-5076 online
DOI: 10.3109/02699206.2012.683672

Computer-assisted analysis of spontaneous speech: quantification of basic parameters in aphasic and unimpaired language

KATJA HUSSMANN[1], MARION GRANDE[1], ELISABETH MEFFERT[1], SWETLANA CHRISTOPH[1], MARTINA PIEFKE[2,3,4], KLAUS WILLMES[5], & WALTER HUBER[1]

[1]*Section Neurological Cognition Research, Department of Neurology, Medical School, RWTH Aachen University, Aachen, Germany,* [2]*Cognitive Neurology, Department of Neurology, RWTH Aachen University, Aachen, Germany,* [3]*Physiological Psychology, Department of Psychology and Sports Sciences, Bielefeld University, Bielefeld, Germany,* [4]*Center of Excellence "Cognitive Interaction Technology" (CITEC), Bielefeld University, Bielefeld, Germany, and* [5]*Section Neuropsychology, Department of Neurology, Medical School, RWTH Aachen University, Aachen, Germany*

(*Received 31 January 2012; revised 13 March 2012; accepted 4 April 2012*)

Abstract
Although generally accepted as an important part of aphasia assessment, detailed analysis of spontaneous speech is rarely carried out in clinical practice mostly due to time limitations. The Aachener Sprachanalyse (ASPA; Aachen Speech Analysis) is a computer-assisted method for the quantitative analysis of German spontaneous speech that allows for a detailed assessment by means of linguistic basic parameters in an acceptable amount of time. It has previously been proven sensitive for monitoring changes over time. In this study, we present data of 52 aphasic participants whose spontaneous speech was analyzed retrospectively before and after an intensive therapy program. The measured changes are evaluated with reference to normative data of 60 non-brain-damaged speakers. Results confirm good sensitivity to document changes over time. Clinical relevance of changes is assessed with reference to critical score ranges derived from the normative data. Findings provide further evidence of the clinical applicability and usefulness of ASPA.

Keywords: *aphasia*, *linguistic analysis*, *normative data*, *clinical application*, *German*

Introduction

The importance of spontaneous speech[1] analysis is widely accepted among both speech therapists and researchers working in the field of aphasia (Prins & Bastiaanse, 2004). Because improvement of spoken language production is usually one of the most important aims of treatment both for the aphasic speaker and the family, analysis of spontaneous speech should be part of diagnostic assessment, treatment planning and evaluation of treatment outcome (see, e.g. Lind, Kristoffersen, Moen, & Simonsen, 2009; Webster, Franklin, & Howard, 2007). This aspect of diagnostics has become even more important since growing attention is spent on aphasic speakers' abilities to participate in daily life activities (World Health Organisation, 2001).

Correspondence: Katja Hussmann, Section Neurological Cognition Research, Department of Neurology, Medical School, RWTH Aachen University, Pauwelsstr. 30, 52074 Aachen, Germany. Tel: +49 241 80 88556. E-mail: khussmann@ukaachen.de

Yet, in spite of this generally accepted necessity of spontaneous speech analysis, it is rarely applied in clinical work (e.g. Lind et al., 2009; Meffert, Hussmann, & Grande, 2009). If at all, spontaneous speech analysis is most often accomplished as part of standardized and well-established diagnostic batteries for aphasia, for example, the Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barresi, 2000) for English and the Aachener Aphasie Test (AAT; Huber, Poeck, & Weniger, 1984; Huber, Poeck, Weniger, & Willmes, 1983) for German. Both aphasia tests comprise qualitative analysis of spontaneous speech by means of rating scales on which the examiner has to rate a speaker's abilities on six scales with respect to different aspects of oral language communication. Although well-established in clinical routine (see, e.g. Katz et al., 2000 for English language batteries) and often used in studies aiming at investigating topics of aphasia (see, e.g. Crepaldi et al., 2011 and Bastiaanse, 2011 for recent application of the AAT procedure for eliciting spontaneous speech in the Italian and Dutch version, respectively), their usefulness is limited when it comes to a more detailed analysis: inter-rater reliability is not perfect; in addition, sensitivity to detect inter-subject differences on particular parameters or to evaluate intra-subject changes over time is often too low – especially for rather subtle changes often occurring in spontaneous speech (Hussmann et al., 2008; Prins & Bastiaanse, 2004).

Alternatively, quantitative methods can be used to analyze spontaneous speech. Here, certain phenomena (basic parameters or aphasic symptoms) are counted exactly and then used to calculate different parameters (Meffert et al., 2009). When presented with detailed user guidelines, such quantitative methods are often superior to rating scales regarding inter-rater reliability (Prins & Bastiaanse, 2004). Additionally, with quantitative methods, it is often possible to detect even subtle changes over time. For English, several different methods have been developed. Examples of well-established, efficient and useful methods for their respective research interests are the quantitative production analysis (Berndt, Wayland, Rochon, Saffran, & Schwartz, 2000), the analysis by means of Text Units by Edwards and Bastiaanse (1998), the analysis of correct information units by Nicholas and Brookshire (1993) and the Shewan spontaneous language analysis system (Shewan, 1988). However, even though at least the latter was explicitly designed to be efficient enough for clinical settings, all these methods are rather time-consuming, be it due to an extensive editing of the speech samples required before analysis or due to an elaborate calculation of parameters after transcription and editing of the language probes. Application within the limited time of clinical practice is therefore not feasible very often.

One more convenient way is to use computer-assisted methods, which facilitate transcription and/or provide calculation of parameters (e.g. Kirmess & Lind, 2010; Kong & Law, 2009). One example of a current computer-assisted method is the Systematic Analysis of Language Transcripts (SALT; Miller & Chapman, 1983). SALT was successfully applied to aphasic language samples (Holland et al., 1985); today, however, it is usually employed in the analysis of child language, for which it had originally been developed. SALT comes with detailed guidelines, and several different versions are available for different research and clinical purposes. Furthermore, comprehensive reference data are available (see, e.g. Heilmann, Miller, & Nockerts, 2010). Regarding the analysis of aphasic language, the AphasiaBank (as aphasia-related part of the TalkBank-Project; see, e.g. MacWhinney, 1995) is a very comprehensive and sophisticated database of language samples which can be transcribed and analyzed by means of different tools (CHAT and CLAN programs). At the moment, the analysis of aphasic language focuses on samples elicited by a story-retelling task (“Cinderella”; see MacWhinney, Fromm, Holland, Forbes, & Wright, 2010). Although available with detailed guidelines, application in time-limited clinical settings is neither the primary focus of the project nor feasible. Even though both former tools cannot be applied in *clinical* analysis of *aphasic* language, they provide comprehensive reference data of control participants (i.e. typically developing children and non-brain-damaged adults, respectively). Although an essential requirement for useful and clinically applicable methods, data of a sufficient number of control participants are often not available for a particular analysis or elicitation method (see, e.g. Webster et al., 2007).

In summary, to meet the requirements both from the investigator's point of view and from the given setting in clinical practice, a method of spontaneous speech analysis has to be easily (i.e. without too much background knowledge) and quickly applicable, should make use of a manageable amount of reasonable (linguistic) parameters, should be sensitive enough to detect subtle changes over time and has to provide clear instruction guidelines and normative data of a sufficient number of control participants (e.g. Lind et al., 2009).

With the Aachener Sprachanalyse (ASPA; Huber, Grande, & Springer, 2005), we have presented a computer-assisted instrument which was developed for the quantitative analysis of German spontaneous speech. Basic parameters on the lexical and syntactic level are counted instead of aphasic symptoms. Hence, the program can be used for the analysis of samples of both impaired and unimpaired language, so that normative data can be collected as a basis for the evaluation of impaired language. In preliminary studies, ASPA was found to be a valid and reliable method (Barthel, Djundja, Meinzer, Rockstroh, & Eulitz, 2006; Grande et al., 2008), also sensitive enough to detect even small changes over time (Grande et al., 2008; Hussmann et al., 2006). Additionally, Meffert et al. (2010) derived normative values of 60 non-brain-damaged speakers as a basis for the comparison with impaired language samples (see Methods section). In that study, the ASPA parameters also proved to be robust against influence by age or educational level. Until now, however, only smaller samples were examined and the evaluation of the quality of the measured changes remained partly unclear (Grande et al., 2008). As a consequence, the aims of this study were (a) to replicate our previous findings regarding the sensitivity to changes over time based on a larger sample and (b) to further investigate the usefulness of ASPA as a clinically applicable instrument for analyzing the spontaneous speech of German aphasic speakers; regarding the clinical applicability, we aimed particularly at (a) investigating the possibility to differentiate fluent and non-fluent aphasic speakers and non-brain-damaged speakers by means of ASPA parameters and at (b) establishing a reasonable evaluation of the measured changes over time referring to their clinical relevance. We focus on the following six parameters: percentage words (W) in contrast to interjections and neologisms, percentage open class words (OCW) in contrast to closed class words, type-token-ratio of open class words (TTR),[2] syntactic completeness (COMPL), complexity (CPX) and mean length of utterances (MLU) in words (see Methods section for details). These basic linguistic parameters were chosen as we expect them to reflect cardinal aphasic symptoms as, for example, word finding difficulties, agrammatism and paragrammatism. Word finding difficulties, for example, frequently lead to production of interjections. Thus, percentage words (W) will decrease in these cases. Agrammatism is characterized, among other things, by truncation of utterances; this will be reflected in low values of syntactic completeness (COMPL). In paragrammatic spontaneous speech, overuse of function words is characteristic, which will lead to a lower percentage open class words (OCW). Thus, we suppose the chosen parameters to represent different types of aphasic spontaneous speech when considered altogether (cf. Grande et al., 2008). Based on our previous findings, we expect the three groups of speakers (non-fluent and fluent aphasic speakers and non-brain-damaged speakers) to differ significantly for the majority of parameters chosen: word finding difficulties and impairment of syntactic structures will supposedly lead to a lower percentage words (W) and less lexical diversity (TTR) in both fluent and non-fluent aphasic speakers in contrast to non-brain-damaged speakers; shorter utterances (MLU) and less complete and less complex utterances (COMPL, CPX) are expected for the non-fluent aphasic speakers, whereas performance of fluent aphasic speakers might be similar to that of the non-brain-damaged participants (especially concerning MLU and CPX). Regarding the percentage open class words (OCW), we expect to find confirmation of the middle position of non-brain-damaged speakers (Meffert et al., 2010) between the fluent aphasic speakers (too low percentage OCW due to paragrammatism) and the non-fluent aphasic speakers (too high due to agrammatism). Concerning the sensitivity to changes over time, we expect to corroborate higher sensitivity of ASPA parameters compared to AAT spontaneous speech rating scales.

Besides, we suggest a method to evaluate the performance of individual aphasic speakers in relation to the spontaneous speech of the non-brain-damaged subjects group and hope to thereby get a better understanding of the clinical relevance of the observed changes.

Methods

Participants

Aphasic speakers. The data of 52 aphasic subjects who had been inpatients on the Aachen Aphasia Ward between 2002 and 2004 were included in this study. All had suffered from a single cerebrovascular accident in the left hemisphere.[3] Descriptive information is given in Table I.

Because treatment on the Aachen Aphasia Ward is an intensive program with at least 8 h per week over 7 weeks, all patients were examined carefully before admission to the ward to ensure their ability to cope with the program and thereby to facilitate a successful outcome. This implies sufficient attention, memory and learning ability. Additionally, all participants had to be native speakers of German. To avoid ceiling or floor effects, patients were excluded when they had scored 0 (very poor) or at least 4 (mild/no impairment) more than once in the semantic, phonological and/or syntactic structure spontaneous speech rating scales of the AAT.[4]

Non-brain-damaged speakers. To obtain normative data as a basis for the evaluation of the aphasic speakers' data, the spontaneous speech of 60 adults (30 females) without language impairment was analyzed in a preceding study (Meffert et al., 2010). Mean age of non-brain-damaged participants was 47.9 years. All were native speakers of German. Exclusion criteria were any neurological or psychiatric illness. They were recruited either while they were inpatients at the Aachen University Hospital (wards for internal or surgical medicine) or via private contacts.

Elicitation of spontaneous speech

Aphasic participants. Spontaneous speech samples of the aphasic subjects were taken from the AAT archive of the Aachen Aphasia Ward and transcribed and analyzed retrospectively using ASPA without knowledge of the AAT scores. The speech samples were – according to the AAT procedure – elicited in the semi-standardized interview covering the topics onset of illness, occupation, family and hobbies (for detailed description of the English version of the AAT, see Miller, Willmes, & De Bleser, 2000). In order to obtain enough analyzable utterances even from non-fluent aphasic

Table I. Descriptive information for the aphasic participants.

Aphasia	Stage	Total	Male	Female	Duration (months) *M* (range)	Age (years) *M* (range)	Severity (MPH) *M* (range)
Fluent	Post-acute	19	15	4	5.1 (0–10)	49.8 (27–68)	50.7 (38.8–58.8)
	Chronic	7	5	2	18.3 (13–36)	51.4 (44–65)	53.7 (43.9–59.1)
	All fluent	26	20	6	8.7 (0–36)	50.2 (27–68)	51.5 (38.8–59.1)
Non-fluent	Post-acute	9	4	5	5.9 (1–11)	52.1 (34–74)	48.1 (38.2–57.0)
	Chronic	17	12	5	34.5 (14–86)	45.1 (22–65)	49.1 (43.0–61.0)
	All non-fluent	26	16	10	24.6 (1–86)	47.5 (22–74)	48.7 (38.2–61.0)
Total		52	36	16	16.6 (0–86)	48.9 (22–74)	50.1 (38.2–61.0)

Notes: Fluent, AAT spontaneous speech syntax-scale score 3 or 4; non-fluent, AAT spontaneous speech syntax-scale score 0, 1 or 2; post-acute: ≤12 months, chronic: >12 months; MPH, mean profile level of AAT subtest performance (T-scores; reliability-weighted average of T-scores per subtest).

speakers, the first 60 clause-like units (CLUs, see below) were transcribed, one CLU being a syntactically or prosodically marked part of spontaneous speech that constitutes a linguistic proposition (see also Appendix 1).

For each participant, spontaneous speech samples from two test administrations were chosen: the most recent AAT diagnostic examination before the intensive treatment (pre-test, t1) and, the AAT administered at the end of the 7 weeks treatment (post-test, t2). If a participant did not produce 60 CLUs – which was often the case for the speakers presenting with non-fluent aphasia – the whole transcript was analyzed. Mean length of transcripts was 51 CLUs at t1 and 52 CLUs at t2. The shortest transcript included in this study contained 12 CLUs (cf. Appendix 2 for results of all participants at t1 individually).

Non-brain-damaged speakers. Spontaneous speech for participants in the group of non-brain-damaged speakers was recorded only once, also following the procedure of the semi-standardized interview of the AAT covering the topics onset of illness, occupation, family and hobbies. Subjects, who were no inpatients at that time, were asked to tell about their last serious illness. In order to obtain transcripts of sufficient length, the first 30 CLUs were transcribed and analyzed.[5] Based on the parameter values, a diagnostic scheme was established defining ranges "average" ($M \pm 1$ SD), "beyond average" ($M \pm$ no more than 2 SD) and "far beyond average" ($M \pm 2$ SD minimum). Table II shows the values for each parameter and its ranges.

Analysis of spontaneous speech

Qualitative analysis (only aphasic participants). At the time of the actual treatment (years 2002–2004, respectively) the speech samples had already been evaluated according to the clinical routine using the six AAT spontaneous speech rating scales "Communicative Behaviour", "Articulation and Prosody", "Formulaic Language", "Semantic Structure", "Phonological Structure" and "Syntactic Structure". On each scale, the examiner has to rate the speaker's performance on a scale from 0 to 5 (0 meaning most severely impaired, 5 meaning unimpaired). For details of the rating scale, see the description of the English AAT (Miller et al., 2000).

Quantitative analysis by ASPA. The ASPA computer program is provided with an instruction manual (Huber et al., 2005) and detailed transcription guidelines (Grande, Springer, & Huber, 2006) of which the most important information is summarized in Appendix 1. High inter-rater reliability (0.90 for segmentation and 0.95 for classification of CLUs, cf. Grande et al., 2008) indicates user-friendly applicability.

Table II. ASPA parameters: value ranges derived from normative data (Meffert et al., 2010).

Parameter (*M*; SD)	Far below average	Below average	Average	Above average	Far above average
W (93.3%; 3.5)	≤86.2	86.3–89.7	89.8–96.8	≥96.9	–
OCW (31.7%; 4.0)	≤23.6	23.7–27.6	27.7–35.7	35.8–39.7	≥39.8
TTR (82.5%; 6.6)	≤69.2	69.3–75.8	75.9–89.1	89.2–95.7	≥95.8
COMPL (76.4%; 11.7)	≤49.9	53–64.6	64.7–88.1	88.2–99.8	≥99.9
CPX (44.3%; 13.8)	≤16.6	16.7–30.4	30.5–58.1	58.2–71.9	≥72
MLU (6.0; 0.9)	≤4.1	4.2–5	5.1–6.9	7–7.8	≥7.9

Notes: W, percentage words; OCW, percentage open class words; TTR, type-token-ratio; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances.

Table III. Description and calculation of relevant ASPA parameters.

Parameter	Calculation
Percentage words (W) compared to interjections and neologisms	(words/(words + interjections + neologisms + unintelligible items)) × 100
Percentage open class words (OCW) compared to closed class words	(open class words/(open class words + closed class words)) × 100
Type-token-ratio of open class words (TTR)	(types OCW/token OCW) × 100
Syntactic completeness (COMPL), i.e. percentage complete clause-like units (CLUs)	(complete CLUs/(complete CLUs + incomplete CLUs)) × 100
Complexity (CPX), i.e. percentage CLUs in compound sentences	(CLUs in compound sentences/(CLUs in compound sentences + simple CLUs)) × 100
Mean length of utterances (MLU) in words	(number of words within CLUs/number CLUs)

ASPA counts the number of items and CLUs and calculates several syntactic and lexical parameters. Based on our previous findings (see Introduction), we focused on the parameters listed in Table III.

Statistical analyses

Comparison and discrimination of groups. To investigate whether the three groups of speakers (non-fluent and fluent aphasic speakers and non-brain-damaged speakers) differ significantly regarding mean basic parameter values, one-way ANOVAs with *post hoc* multiple pairwise comparisons were carried out for each of the six parameters. Potential differences between fluent and non-fluent aphasic speakers regarding sex, age (“younger”: ≤50 years/“older”: >50 years), education (“low”: no academic degree/“high”: academic degree),[6] time interval between t1 and t2 (“short”: <26 weeks/“long”: ≥26 weeks), overall severity of language impairment (measured by the AAT mean profile level, MPH; “mild–moderate”: $\mathrm{MPH} \geq 51$, “moderate–severe”: $\mathrm{MPH} < 51$) and stage of aphasia (“postacute”: ≤12 months, “chronic”: >12 months) were assessed using *t*-tests or Fisher's exact test, respectively.

Significant mean differences between fluent and non-fluent aphasic speakers were found for duration ($\mathrm{t}(31) = 3.53, p = 0.001$) and overall severity of aphasia ($\mathrm{t}(50) = -2.005, p = 0.05$); these two variables were introduced as covariates in the group comparisons to prevent any confounding influence: analyses of covariance (ANCOVA) were carried out for each of the six ASPA parameters to investigate the influence of the variables type of aphasia, age, time between t1 and t2, education and sex.

Finally, stepwise discriminant analyses were carried out to go beyond mean group differences and study the possibility to differentiate between impaired and unimpaired language samples by means of the ASPA parameters.

Change over time. To determine significant change over time for individual participants, a critical difference for each ASPA parameter was calculated using Cronbach's alpha[7]. Changes from t1 to t2 are considered statistically significant for an individual participant, if the difference between t1 and t2 for an ASPA parameter exceeds the critical difference[8] for this parameter. Critical differences for all parameters are shown in Table IV.

Regarding the AAT rating scales, observed changes are considered substantial if they are larger than to be expected due to interrater variability; for all AAT spontaneous rating scales the critical difference is two scores. Regarding sensitivity to change, the number of significant changes

Table IV. ASPA parameters: critical differences.

ASPA parameter	Critical difference
W	6.75
OCW	10.71
TTR	13.19
COMPL	14.76
CPX	15.08
MLU	0.99

Notes: W, percentage words; OCW, percentage open class words; TTR, type-token-ratio; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances. Critical differences for W, OCW, TTR, COMPL and CPX are given as percentage points, for MLU in words.

between t1 and t2 as assessed by the AAT spontaneous speech rating scales *vs*. by the ASPA parameters were counted, respectively.

Quality and clinical relevance of changes. For an analysis of the clinical relevance of the observed changes after treatment, the individual participant's ASPA parameter values were related to the normative value ranges derived from non-brain-damaged speakers. A significant change (i.e. score difference equal to or larger than the critical difference) resulting in a change of category in the scheme comprising five categories from far below average to far above average (derived from the normative data; cf. Table II and Appendix 3) was considered to indicate a definite, clinically relevant improvement or deterioration, respectively. Significant changes not leading to a category change were taken to indicate a trend. Thereby, the operational definition of improvement vs. deterioration follows – for the time being – the rule “the more the better” for all parameters except for the percentage open class words (OCW). We are aware that very long utterances (MLU) and very complex sentence constructions (CPX) would be no reasonable treatment goals. However, so far, the normative data do not supply sufficient information for valid cut-off scores (Meffert et al., 2010). In contrast to most parameters, the following definition for the percentage open class words takes into account the central position of non-brain-damaged speakers between the fluent and the non-fluent aphasic group: improvement is present for a change from too high (=“agrammatic”) as well as from too low (=“paragrammatic”) percentage open class words toward the average value range. Statistical analyses were carried out by means of SPSS for Windows.

Results

Comparison and separation of groups

Significant differences between all groups were present for the four parameters percentage words, percentage open class words, syntactic completeness and complexity. Regarding the type-token-ratio of open class words (TTR), the two aphasic groups did not differ significantly while they both showed a significantly lower mean TTR than the group of non-brain-damaged speakers. Non-fluent speakers produced shorter utterances on average (i.e. lower MLU) than both fluent aphasic speakers and non-brain-damaged speakers (cf. Table V).

The additional ANCOVAs revealed only an effect of age for the mean length of utterances: older speakers produced significantly longer utterances ($M = 5.0$, SD $= 1.4$) than younger speakers ($M = 4.1$, SD $= 1.6$; $F(1,50) = 4.52$, $p = 0.038$). The factors time interval (between t1 and t2), education and sex had no significant effect.

Table V. ANOVA results for group comparisons of ASPA parameters before therapy (t1).

Parameter (t1)	Non-fluent	Fluent	Non-brain-damaged	*F*-test	*p*	*Post hoc* tests[b]
W	58.2 (14.6)	84.1 (8.4)	93.3 (3.5)	$F(2,42) = 100.38^a$	<.001	(NFL) < (FL) < (NS)
OCW	42.5 (15.3)	24.9 (5.6)	31.7 (4.0)	$F(2,33) = 19.44^a$	<.001	(NFL) > (NS) > (FL)
TTR	62.8 (14.2)	62.6 (8.1)	82.5 (6.6)	$F(2,47) = 48.71^a$	<.001	(NFL, FL) < (NS)
COMPL	15.4 (13.5)	55.3 (18.3)	74.3 (12.5)	$\mathrm{F}(2,60) = 146.29^a$	<.001	(NFL) < (FL) < (NS)
CPX	4.0 (8.9)	31.6 (16.7)	44.2 (13.6)	$F(2,63) = 83.39^a$	<.001	(NFL) < (FL) < (NS)
MLU	3.2 (0.9)	5.8 (0.9)	6.0 (0.9)	$F(2,109) = 84.99$	<.001	(NFL) < (FL, NS)

Notes: FL, fluent; NFL, non-fluent; NS, non-brain-damaged speaker.
[a] Due to variance heterogeneity (Levene test $p < 0.10$) correction of *df* (Brown–Forsythe).
[b] With type I error-adjustment; Tukey HSD/Tamhane test in case of homogeneous/heterogeneous variances.

Table VI. Stepwise linear discriminant analysis for all three groups: re-classification results in absolute numbers (and percent).

	Predicted group membership			
Group	Non-fluent	Fluent	Non-brain-damaged	Total
Non-fluent	**22 (84.6)**	4 (15.4)	0 (0)	26 (100)
Fluent	2 (7.7)	**22 (84.6)**	2 (7.7)	26 (100)
Non-brain-damaged	0 (0)	0 (0)	**60 (100)**	60 (100)

Notes: Variables included: W, TTR, OCW, COMPL. The correct re-classifications (leaving-one-out method for cross-validation) for each group in bold; overall re-classification 92.9%.

Table VII. Stepwise linear discriminant analysis for differentiating between non-fluent and fluent aphasic speakers: re-classification results in absolute numbers (and percent).

	Predicted group membership		
Group	Non-fluent	Fluent	Total
Non-fluent	**25 (96.2)**	1 (3.8)	26 (100)
Fluent	2 (7.7)	**24 (92.3)**	26 (100)

Notes: Variables included: MLU, COMPL. The correct re-classifications (leaving-one-out method for cross-validation) for each group in bold; overall re-classification 94.2%.

In order to investigate how well the groups of speakers could be separated by means of the ASPA parameters, stepwise linear discriminant analyses were carried out for (a) all three groups and (b) the non-fluent and fluent aphasic speakers only. In all discriminant analyses, more than 90% of cross-validated grouped cases were correctly re-classified. The variables which contributed significantly to this distinction between all three groups were syntactic completeness, type-token-ratio, percentage words and percentage open class words, whereas mean length of utterances and syntactic completeness were sufficient to discriminate the two aphasic groups (cf. Tables VI and VII).

Change over time

In order to evaluate the sensitivity to change over time after treatment, the number of participants with at least one significant score change from t1 to t2 both for ASPA parameters and AAT rating scales was counted. The results are given in Table VIII.

Table VIII. Number of participants with at least one significant change from t1 to t2 for ASPA parameters and/or the AAT spontaneous speech rating scales.

	ASPA parameters		
AAT rating scales	Change	No change	Total
Change	6	1	7
No change	37	8	45
Total	43	9	52

Table IX. Number of participants with at least one clinically relevant change in ASPA parameters.

	Improvement			Deterioration	
	Definite	Trend		Definite	Trend
	15	5		8	6
			9[a]		
Total			43[b]		

Note: Definite/trend, significant change resulting in category change (cf. Table II or Appendix 3) yes/no.
[a] Participants with at least one improvement and at least one deterioration.
[b] Eight participants with no change of spontaneous speech at all, 1 with change only measured by the AAT rating scales (cf. Table VIII).

The table clearly demonstrates greater sensitivity of ASPA parameters to indicate change in performance (binomial test, $p < 0.001$).

Quality and clinical relevance of changes

Applying the evaluation scheme for clinically relevant observed changes to the participants in Table VIII led to the results reported in Table IX.

Discussion

Our study aimed on the one hand at supporting previous findings concerning the spontaneous speech performance of non-fluent and fluent aphasic as well as non-brain-damaged speakers measured by ASPA parameters based on larger samples. On the other hand, we wanted to further investigate the clinical applicability and usefulness of ASPA as a method of spontaneous speech analysis, in particular, the possibility to differentiate between the groups of aphasic and non-brain-damaged speakers by means of ASPA parameters and to document clinical relevance of measured changes over time. These aspects will be discussed in detail below.

Comparison of groups

We expected non-fluent aphasic speakers to differ significantly from non-brain-damaged speakers in all ASPA parameters, whereas fluent aphasic speakers might reach the mean length of utterances and syntactic complexity levels of the non-brain-damaged participants. These expectations were corroborated to a large extent. As expected, the non-fluent speakers differed significantly for all parameters from the non-brain-damaged participants, their spontaneous speech characterized by more interjections (in contrast to words), a high percentage open class words, shorter and less complete as well as less complex utterances compared with non-brain-damaged subjects. Also, lexical diversity is lower than that of the non-brain-damaged participants. This pattern is clearly in accordance with the clinical impression of non-fluent speakers. Thus, the basic parameters seem appropriate to capture salient properties of spontaneous speech in non-fluent aphasia. With the exception of the type-token-ratio, non-fluent speakers' performance differs significantly from fluent speakers' indicating usefulness of the ASPA parameters when having to compare the two aphasic groups. Also, the comparison of fluent speakers with non-brain-damaged speakers yielded reasonable findings: fluent speakers produced more interjections (in contrast to words), showed a higher percentage open class words, a lower type-token-ratio and fewer complete utterances. Not unexpectedly, fluent aphasic speakers produced utterances of similar mean length compared to non-brain-damaged subjects. The former group did not, however, achieve a similar level of complexity as we had assumed on the basis of our previous findings. Larger sample sizes might reveal that – even if there are fluent participants who actually produce many complex utterances – there is a large number of fluent speakers whose grammatical deviations do not become manifest in an increased tendency to produce subordinate structures (see, e.g. Edwards & Bastiaanse, 1998).

The non-brain-damaged speakers show a higher type-token-ratio (TTR) than the non-fluent and fluent aphasic speakers; but the two aphasic groups do not differ significantly. In fact, they achieved almost equal means. As the respective TTR ranges of the non-fluent and fluent groups vary strongly from 36.8 to 95 (extremely high values due to extremely short transcripts with few, but different words) and 48.9 to 80.3, respectively, it becomes obvious that the different lengths of transcripts led to a cancellation effect, as reported in the literature (Wright, Silverman, & Newhoff, 2003). In studies which aim at the analysis of speech samples from both fluent and non-fluent speakers, the type-token-ratio is a problematic parameter, because one will probably always have to accept shorter transcripts from non-fluent speakers (see, e.g. Crepaldi et al., 2010; Mayer & Murray, 2003). Hence, in order to assess lexical diversity, it is advisable to draw on one of the alternative measures (e.g. Wright et al., 2003), even if calculation is more complex or more specific knowledge is needed for analysis. Alternatively, one can choose another reference instead of transcript length. Bastiaanse and Jonkers (1998) investigated verb diversity in spontaneous speech of agrammatic speakers by means of transcripts controlled for number of words instead of, for example, CLUs. Results revealed, however, that not the number of words, but of tokens instead (in their case verbs) was relevant for reasonable assessment of the type-token-ratio. Thus, the type-token-ratio can be calculated even with different sample sizes, as long as the number of tokens is controlled; results of Bastiaanse and Jonkers would suggest differentiating between type-token-ratios of different word classes rather than calculating one type-token-ratio of all open class words taken together (Bastiaanse & Jonkers, 1998).

The ANCOVAs revealed no influence of education, sex or time of interval between t1 and t2 for the aphasic groups. Rather unexpectedly, an effect of age emerged for the parameter mean length of utterances: the younger speakers produced significantly shorter utterances than the older ones. There is no straightforward explanation, especially as this finding is in contrast to the main effect of age that was found in the normative data (Meffert et al., 2010); here, older speakers produced shorter utterances than younger speakers. Influence of aphasia seems to dominate the influence of age. More data – larger groups of fluent and non-fluent aphasic speakers from the different age groups – are surely needed before an explanation is possible.

Differentiating the three groups by means of the ASPA parameters

The discriminant analyses on the ASPA parameters yielded very high proportions of correctly re-assigned cases. Differentiating all three groups was successful by means of the parameters syntactic completeness, type-token-ratio, percentage words and percentage open class words and resulted in an overall (cross-validated) correct re-classification of 92.9% of all cases.

Eight cases were not correctly allocated: four non-fluent speakers were re-classified as fluent (participants 29, 32, 42 and 48; see Appendix 2 for an overview of the individual parameter values at t1 of all participants), two fluent speakers were re-classified as non-fluent (participants 12 and 16) and two fluent speakers were re-classified as non-brain-damaged speakers (participants 10 and 24). A closer look at the patterns of parameter values of the incorrectly classified non-fluent speakers reveals one common feature: they all show a percentage open class words (OCW) that does not match the expected high percentage open class words in the characteristic non-fluent pattern. One of the participants (32) in contrast shows an extremely low percentage open class words which would rather correspond to a fluent pattern. Because percentage open class words is one of the parameters included in the discriminant analysis, this may explain the incorrect re-classification in this case.

The two fluent speakers who were incorrectly re-classified as non-brain-damaged speakers (10 and 24) present with the following pattern: with one exception the value range is “average” or “below average” with only little margin to the “average” range for all parameters. Applying Crawford's single-case statistic program[9] to compare the values of these two participants with the normative data, one participant's (10) performance shows no significant difference to the values of the non-brain-damaged speakers in any parameter (one-tailed probability, $\alpha = 5\%$). The second participant (24) shows only one significant difference (percentage open class words, $p < 0.001$). Thus, re-classification as non-brain-damaged speakers is not surprising, and these two participants clearly represent speakers, for whom further parameters would be needed to differentiate them from the non-brain-damaged speakers.

The two fluent participants who were re-classified as non-fluent speakers (12 and 16) show a parallel pattern, namely values “far below average” for all parameters included in the analysis, which means fewer complete utterances and a lower percentage of words than expected in comparison to fluent speakers.

Regarding the discriminant analysis for the two aphasic groups, a similarly high percentage of correct re-classifications was achieved (94.2%). Only the two parameters mean length of utterances and syntactic completeness were required for group separation. Two participants presenting with fluent aphasia were incorrectly re-classified as non-fluent (participants 3 and 12), whereas only one non-fluent aphasic speaker was incorrectly re-classified (participant 32). This latter participant was also falsely re-classified as fluent in the discriminant analysis for all three groups; she showed a mean length of utterances in the “average” range, which was most probably the reason for the repeated wrong reassignment. Also, one of the two incorrectly re-classified fluent speakers (12) had already been falsely reassigned in the first analysis due to his performance “far below average” in all parameters. Ranges “far below average” and “below average” for syntactic completeness and mean length of utterances, respectively, were also shown by the other fluent speaker re-classified as non-fluent.

Quality and clinical relevance of changes

As expected, we could find corroborating evidence for the higher sensitivity to change of the ASPA parameters in comparison to the AAT spontaneous speech rating scales.

Aiming at clinical feasibility, we suggested an evaluation scheme based on five categories derived from the normative data. Applying this scheme, the measured changes can be clearly identified as improvement or deterioration of performance (or a trend toward one direction) in 34 of the 52 participants. Fifteen participants showed an improvement from t1 to t2 measured by a significant change in at least one of the ASPA parameters that led to a concomitant change of category. Eleven of these 15 participants also achieved significant improvement either in one AAT subtest or for one of the AAT spontaneous speech rating scales. For those eleven participants, we diagnosed general improvement of aphasia (presumably induced by the intensive treatment program) that even becomes manifest in spontaneous speech.

Five participants showed a trend to improvement according to our evaluation scheme, that is, significant change for at least one ASPA parameter, but no change of category. Three of these participants did not show significant change either in the AAT subtests or the AAT spontaneous speech rating scales. All three presented with non-fluent aphasia and had overall low values for the ASPA parameters. Significant changes concerned percentage words and mean length of utterances. Our assumption here is that the observed improvement is substantial but – given the low initial values – further progress is needed before average levels can be reached. One of the five participants with a trend to improvement also showed significant improvement in the AAT subtests Token Test and Written Language; improvement measured by ASPA concerned complexity (CPX), the CPX value, however, being in the "average" range already at t1. His improvement may therefore be mostly due to improvement in more complex structures that were possibly practiced in treatment of written language. The fifth participant showing a trend to improvement had an increase of percentage words, while simultaneously his naming performance, measured by the AAT subtest "Naming", decreased significantly. This result is surely against our expectations as we would assume an increase of percentage words (meaning less interjections/neologisms) to be caused by less word finding difficulties or better naming abilities. However, the relation between performance in confrontational naming tests and word retrieval in connected speech is not yet resolved. Whereas Mayer and Murray (2003) did not find significant correlations, Herbert, Hickin, Howard, Osborne, and Best (2007) found evidence of a quantitative relation between word retrieval in picture naming and in conversation.

The performance of 14 participants decreased between t1 and t2 – a development clearly unwanted and rather unexpected in view of the fact that all these participants had intensive treatment. When looking at the decreased parameters in question no particular parameter shows up more often than the other; also, there were participants who showed improvement and participants who showed deterioration of performance in the AAT along with the deterioration of performance measured by ASPA. The only pattern that emerged in four participants (once definitely clinically relevant deterioration, three times a trend) was an increase of percentage open class words (=deterioration!) associated with improved naming performance in the AAT subtest. If one assumes that there is a relation between confrontation naming ability and spontaneous speech, then a higher percentage open class words can be explained by improved word retrieval. Because all three participants presented with non-fluent aphasia, they all showed already a percentage open class words above average, which was further increased by using treatment-induced reactivated words in spontaneous speech.

The patterns of change of nine further participants have to be left unexplained. These nine participants showed improvement *and* deterioration at the same time, as measured by at least one ASPA parameter, respectively. Again, each parameter occurred, but no combination of parameters stood out. Additional data may reveal an association between complexity and type-token-ratio, the two parameters showing significant change in the performance of five participants (with both parameters sometimes being improved, sometimes being deteriorated, respectively). However, the type-token-ratio has to be considered with caution due to methodological limitations (see below). This pattern of change clearly has to be corroborated by other measures. Not too surprising is the fact that there were eight participants showing no change of spontaneous speech at all, given that transfer of improved performance into spontaneous speech is not always achieved. Finally, one participant showed improvement only in the AAT spontaneous speech rating scale "Phonological structure" and in the AAT subtest "Repetition". As there is no ASPA parameter that covers phonological structure, it is comprehensible that a change at that level alone cannot be detected by the ASPA parameters used.

Normative data in clinical practice

Some limiting aspects notwithstanding (see below), we consider our findings as evidence for both the clinical usefulness of the ASPA parameters and the importance of normative data as a basis of comparison between impaired and unimpaired performance. Analysis of spontaneous speech by means of ASPA in combination with normative data provides the investigator with the possibility of classifying a participant's performance in comparison to non-brain-damaged control participants. Furthermore, normative data help to define treatment goals in a transparent way. Additionally, change over time can be evaluated for significance via the calculated critical differences. As an example of application in clinical practice, the parameter values determined for one participant (# 16) are given in Table X. He achieved one clinically relevant improvement (CPX) and two trends of improvement (W, COMPL). TTR remained constant far below average. For OCW and MLU, a change into another value range occurred, leading to an overall better spontaneous speech profile at t2. Because the differences in OCW and MLU values did not exceed the respective critical differences, these numerical changes are not marked as statistically significant.

Several aspects limiting the generalizability of our approach need to be mentioned. At the moment, the findings refer only to spontaneous speech for the German language, elicited via the

Table X. Diagnostic scheme presenting parameter values of participant 16 at t1 and t2 with regard to the normative value ranges (from Meffert et al., 2010).

Patient: participant 16	t1: 14 April 2003		t2: 27 May 2003		
	48 CLUs. 466 words		60 CLUs. 518 words		
Duration of aphasia	4 weeks		10.5 weeks		
Parameter (mean; SD; crit. diff.)	Far below avg.	Below avg.	Average	Above avg.	Far above avg.
W (93.3%; 3.5; 6.75)	≤86.2	86.3–89.7	89.8–96.8	≥96.9	–
t1	62.9				
t2	82.2[a]				
OCW (31.7%; 4.0; 10.71)	≤23.6	23.7–27.6	27.7–35.7	35.8–39.7	≥39.8
t1	18				
t2		23.7			
TTR (82.5%; 6.6; 13.19)	≤69.2	69.3–75.8	75.9–89.1	89.2–95.7	≥95.8
t1	51.2				
t2	55.3				
COMPL (76.4%; 11.7; 14.76)	≤49.9	53–64.6	64.7–88.1	88.2–99.8	≥99.9
t1	14.6				
t2	47.5[a]				
CPX (44.3%; 13.8; 15.08)	≤16.6	16.7–30.4	30.5–58.1	58.2–71.9	≥72
t1	12.5				
t2			30.5[b]		
MLU (6.0; 0.9; 0.99)	≤4.1	4.2–5	5.1–6.9	7–7.8	≥7.9
t1				7.6	
t2					8.4

Notes: W, percentage words; OCW, percentage open class words; TTR, type-token-ratio; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances. Critical differences for W, OCW, TTR, COMPL, and CPX are given as percentage points, for MLU in words. Value ranges derived from normative data (Meffert et al., 2010): average (mean ± 1 SD); below/above avg., below/above average (mean ± not more than 2 SD); far below/above avg., far below/above average (mean ± 2 SD minimum); CLU, clause-like unit; t1/2, first/second time of assessment.

[a] Trend of change.

[b] Clinically relevant change.

semi-standardized interview of the AAT. However, we are confident that these simple basic parameters will be applicable even in other languages. The elicitation method used with the AAT is clinically very feasible since the questions are asked anyway for documentation of the patient's case history. With other elicitation procedures and other native languages further normative data will have to be acquired, because it has been shown that different elicitation methods may result in different parameter profiles (see, e.g. Sproedefeld et al., 2008). Second, as in many clinical studies comprising fluent and non-fluent aphasic speakers, some transcripts are very short due to the limited productions of the non-fluent speakers. We tried to counter this problem by extending the length of the aphasic speakers' transcripts to 60 CLUs (in contrast to 30 CLUs for control subjects which yielded 183 words on average); but this did not lead to more analyzable speech in all cases. Third, the type-token-ratio is known to be a parameter sensitive to the length of transcripts (see above). Not unexpectedly, our results were ambiguous. Therefore, we would suggest choosing a different method for analyzing lexical diversity, for example, by using transcripts with a comparable number of open class words or even nouns or verbs instead of controlling for the total number of CLUs (cf. Bastiaanse & Jonkers, 1998). Finally, because no follow-up testing was available, the long-term stability of spontaneous speech performance could not be evaluated.

Conclusion

The mentioned limitations notwithstanding, we consider the results to further support usefulness of ASPA as a method of analysis of spontaneous speech for several reasons. The two aphasic groups differed significantly in all but one parameter. Discriminant analyses yielded overall (cross-validated) correct re-classification of more than 90% of all cases. Analysis by ASPA parameters revealed significantly more changes over time than analysis by the AAT spontaneous speech rating scales. Finally, quality and clinical relevance of change could be clearly demonstrated in 34 of the 52 participants. Thus, the basic parameters computed allow differentiating between impaired and unimpaired language, and they are sensitive enough to detect even subtle changes over time. Clinical relevance of parameter changes can be judged with reference to normative data. Regarding clinical applicability, high inter-rater reliability indicates feasibility even in time-limited settings. Thus, ASPA provides both the researcher and the speech therapist with a reasonable method for the quantitative analysis of spontaneous speech.

Acknowledgments

This research was supported by the START program of the Faculty of Medicine, RWTH Aachen University, Aachen. Some of the results were presented as work in progress at the “7.Jahrestagung der Gesellschaft für Aphasieforschung und –behandlung” (poster presentation and corresponding short printed report, see Hussmann et al., 2008). We thank Roelien Bastiaanse and one anonymous reviewer for helpful comments and suggestions for improvement.

Declaration of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes

1. We regard the terms *spontaneous language* and *spontaneous speech* as equivalent in the context of this study; we use the latter term following the definition of Prins and Bastiaanse (2004).
2. The type-token-ratio is known to be a parameter sensitive to different length of transcripts (Wright, Silverman, & Newhoff, 2003); but as it is also a well-established, easily administered parameter indicating lexical richness, we decided to include it tentatively in this study being aware of the required caution when interpreting the results.

3. One patient had suffered from a cerebro-vascular accident in the right hemisphere. As he was ambidextrous and presented with typical aphasic symptoms, he was included in this study.
4. Score 0 in more than one of these rating scales would characterize most severe language impairment with few or even no analyzable utterances so that the speech sample would not be long enough to be analyzed at all. A score of 4 more than once would indicate only mild to very mild impairment. Whether the ASPA parameters are sensitive enough to measure changes in such mild impairments will be the subject of further studies.
5. Transcription of 30 CLUs led to 183 words (open and closed class) on average which we considered as sufficient (acceptable length in manageable amount of time).
6. As the data were analyzed retrospectively and no detailed information about the participants' educational level was present, their respective classification to the high- or low-level group was concluded from the participants' last occupation before onset of illness.
7. Critical differences were calculated according to the formula $1.645 \times \text{SD} \times \sqrt{(2 \times (1 - \text{Cronbach's alpha}))}$. For details of the procedure, see Grande et al. (2008).
8. Type I error level of 10% as commonly used in individual diagnostic decisions.
9. SINGLIMS.EXE is a statistical program for comparison of an individual's score on a single test with the score of a normative or control sample; download: http://www.abdn.ac.uk/~psy086/dept/psychom.htm#conflims.

References

Barthel, G., Djundja, D., Meinzer, M., Rockstroh, B., & Eulitz, C. (2006). Aachener Sprachanalyse (ASPA): Evaluation bei Patienten mit chronischer Aphasie. *Sprache Stimme Gehör*, *30*, 103–110.

Bastiaanse, R. (2011). The retrieval and inflection of verbs in the spontaneous speech of fluent aphasic speakers. *Journal of Neurolinguistics*, *24*, 163–172.

Bastiaanse, R., & Jonkers, R. (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology*, *12*, 951–969.

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *QPA – quantitative production analysis*. Hove: Psychology Press.

Crepaldi, D., Ingignoli, C., Verga, R., Contardi, A., Semenza, C., & Luzzatti, C. (2010). On nouns, verbs, lexemes, and lemmas: Evidence from the spontaneous speech of seven aphasic patients. *Aphasiology*, *25*, 71–92.

Edwards, S., & Bastiaanse, R. (1998). Diversity in the lexical and syntactic abilities of fluent aphasic speakers. *Aphasiology*, *12*, 99–117.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *The assessment of aphasia and related disorders* (3rd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Grande, M., Hussmann, K., Bay, E., Christoph, S., Piefke, M., Willmes, K., & Huber, W. (2008). Basic parameters of spontaneous speech as a sensitive method for measuring change during the course of aphasia. *International Journal of Language and Communication Disorders*, *43*, 408–426.

Grande, M., Springer, L., & Huber, W. (2006). Richtlinien für die Transkription mit dem Programm ASPA (Aachener Sprachanalyse). *Sprache Stimme Gehör*, *30*, 179–185.

Heilmann, J., Miller, J., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools*, *41*, 84–95.

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2007). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, *22*, 184–203.

Holland, A. L., Miller, J., Reinmuth, O. M., Bartlett, C., Fromm, D., Pashek, G., Stein, D., & Swindell, C. (1985). Rapid recovery from aphasia: A detailed language analysis. *Brain and Language*, *24*, 156–173.

Huber, W., Grande, M., & Springer, L. (2005). *Aachener Sprachanalyse – Handanweisung*. Aachen: Delta Systems.

Huber, W., Poeck, K., & Weniger, D. (1984). The Aachen aphasia test. In F. C. Rose (Ed.), *Progress in aphasiology* (pp. 291–303). New York: Raven Press.

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). *Aachener Aphasie Test*. Göttingen: Hogrefe.

Hussmann, K., Grande, M., Bay, E., Christoph, S., Springer, L., Piefke, M., & Huber, W. (2006). Aachener Sprachanalyse (ASPA): Computergestützte Analyse von Spontansprache anhand von linguistischen Basisparametern. *Sprache Stimme Gehör*, *30*, 95–102.

Hussmann, K., Grande, M., Bay, E., Christoph, S., Willmes, K., & Huber, W. (2008). Quantitative Analyse anhand linguistischer Basisparameter zur Status- und Verlaufsdiagnostik aphasischer Spontansprache. *Bulletin Aphasie und verwandte Gebiete*, *23*, 23–35.

Katz, R. C., Hallowell, B., Code, C., Armstrong, E., Roberts, P., Pound, C., & Katz, L. (2000). *International Journal of Language and Communication Disorders*, *35*, 303–314.

Kirmess, M., & Lind, M. (2010). Spoken language production as outcome measurement following constraint induced language therapy. *Aphasiology*, *25*, 1207–1238.

Kong, A. P. H., & Law, S. P. (2009). A linguistic communication measure for monitoring changes in Chinese aphasic narrative production. *Clinical Linguistics & Phonetics*, *23*, 255–269.

Lind, M., Kristoffersen, K. E., Moen, I., & Simonsen, H. G. (2009). Semi-spontaneous oral text production: Measurements in clinical practice. *Clinical Phonetics & Linguistics*, *23*, 872–886.

MacWhinney, B. (1995). *The CHILDES-project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, *24*, 856–868.

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology*, *17*, 481–497.

Meffert, E., Grande, M., Hussmann, K., Christoph, S., Willmes, K., & Huber, W. (2010). Basisparameter ungestörter Spontansprache: Voraussetzung für die Aphasiediagnostik. *Sprache Stimme Gehör*, *34*, 24–34.

Meffert, E., Hussmann, K., & Grande, M. (2009). Linguistic analysis of spontaneous language in aphasia. In K. Alter, M. Horne, M. Lindgren, M. Roll, & J. von Koss Torkildsen (Eds.), *Brain talk: Discourse with and in the brain* (pp. 321–329). Lund: Lunds Universitet.

Miller, J., & Chapman, R. (1983). Systematic analysis of language transcripts (SALT): A computer program designed to analyze free speech samples. Madison: University of Wisconsin–Madison, Waisman Center, Language Analysis Laboratory.

Miller, N., Willmes, K., & De Bleser, R. (2000). The psychometric properties of the English version of the Aachen Aphasia Test (EAAT). *Aphasiology*, *14*, 683–722.

Nicholas, L.E., & Brookshire, R.H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, *36*, 338–350

Prins, R., & Bastiaanse, R. (2004). Review: Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18*, 1075–1091.

Shewan, C. M. (1988). The Shewan Spontaneous Language Analysis (SSLA) System for aphasic adults: Description, reliability, and validity. *Journal of Communication Disorders*, *21*, 103–138.

Sproedefeld, A., Hussmann, K., Grande, M., Bay, E., Christoph, S., Willmes, K., & Huber, W. (2008). Der Einfluss der Erhebungsmethode auf die spontansprachliche Leistung bei Aphasie. *Bulletin Aphasie und verwandte Gebiete*, *23*, 7–21.

Webster, J., Franklin, S., & Howard, D. (2007). An analysis of thematic and phrasal structure in people with aphasia: What more can we learn from the story of Cinderella? *Journal of Neurolinguistics*, *20*, 363–394.

World Health Organisation. (2001). International classification of functioning, disability and health. Retrieved from http://www.who.int/classifications/icf/en/

Wright, H. H., Silverman, S. W., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, *17*, 443–452.

Appendix 1. ASPA: Guidelines for Transcription and General Guidelines – Summary

The most important points of the guidelines are summarized to indicate our general approach. For details, especially regarding the particular requirements for using the computer program, see Huber et al. (2005) and Grande et al. (2006, 2008).
General guidelines:

- The speech sample is transcribed orthographically. Phonemic deviations are transcribed as precisely as possible; if the target word is recognizable, it has to be added in round brackets. Unintelligible utterances are marked by XXX.
- Pauses are indicated by full stops (short pause = 1–2 s: one full stop; medium–length pause = 2–5 s: two full stops; long pause = > 5 s: three full stops (e.g. “then … um”).
- Curly brackets denote comments regarding the speaker's non-verbal behavior (e.g. {patient points at his arm}) or deviations in articulation, etc. (e.g. {whispers}).
- The transcript should contain at least 30 CLUs (see below for a definition).

Editing of the transcript:

- CLUs have to be identified and marked by square brackets. A CLU is defined as a syntactically and/or prosodically marked portion of spontaneous speech referring to a proposition. A proposition is usually present when at least one open class word is given (e.g. “eh [munich] … [stroke]”). Turns containing only interjections and/or neologistic or unintelligible utterances do not contain a CLU (e.g. “he .. um . XXX no … and then”)
- Criteria for segmentation of the speech sample into CLUs are: (1) syntactic structure, (2) intonation, (3) pauses and (4) semantic structure. These criteria have to be applied in order of priority, that is, if syntactic structure is present and recognizable, division into CLUs follows the syntactic structure. If the syntactic structure alone does not give enough information, intonation has to be considered as an additional indication of possible propositional units (e.g. falling intonation = end of unit). Accordingly, pauses and semantic structure are taken into account additionally, if necessary.
- Brackets are put hierarchically to indicate syntactic complexity, that is, brackets around subordinate CLUs are placed within the brackets of the main CLU (e.g. “[I came back [because I felt sick]]”).
- Interjections may occur inside CLUs or outside; here, intonation and pauses are the decisive criteria (e.g. “um um … [it is not easy]”, “[and then I did not know um see it]”).
- All CLUs have to be classified as complete (C), incomplete (I) or elliptic (E). A CLU is complete if a verb and all its obligatory verb arguments are given. Wrong or missing inflection morphemes or function words do not lead to incompleteness (e.g. “[I was hospital](C)”). Main CLU and subordinate CLU are classified separately regarding completeness. Only the first elliptic utterance produced to answer a preceding open question is defined as ellipsis here (e.g. “how are you?” “[fine](E)”).
- Word class: The program assigns word class to each item based on an internal lexicon. Items which are not included in the lexicon are classified as open class. Therefore, neologisms have to be reclassified by the investigator. Open class words are nouns, verbs, adjectives and adverbs derived from these categories. Closed class words are auxiliaries, determiners, pronouns, conjunctions and genuine adverbs. The particles “yes” and “no” are defined as interjections.
- For correct calculation of TTR citation forms of all inflected words have to be added in round brackets by the investigator.

Appendix 2. Demographic and Clinical Data of All Aphasic Speakers at t1

Part.	Sex	Age (yrs)	Dur. (mths)	Severity (T-score)	t1–t2 (weeks)	Type	Total CLUs	Total words	W	OCW	TTR	COMPL	CPX	MLU
1	m	36	5	50.69	36.1	FL	60	283	89.3	23.7	62.7	63.8	51.7	4.7
2	m	51	9	55.96	65.0	FL	60	339	83.3	25.4	54.7	60	21.7	5.6
3	m	62	1	45.85	3.9	FL	60	275	78.8	12.7	54.3	15	3.3	4.4
4	m	46	10	54.02	23.9	FL	60	320	86.5	30.3	63.9	60	23.3	5.1
5	f	48	15	59.04	97.3	FL	60	340	71.3	24.1	68.3	39.7	17.2	5.4
6	f	43	5	48.96	60.0	FL	60	385	87.9	23.4	48.9	64.4	33.9	6.3
7	m	68	10	58.77	6.0	FL	60	297	84.4	27.6	73.2	53.3	13.3	4.8
8	m	43	0	57.66	7.1	FL	60	373	92.3	30	62.5	78.3	53.3	6.1
9	f	44	18	54.57	25.4	FL	60	349	93.6	30.4	57.5	63.3	36.7	5.8
10	m	55	9	53.53	33.9	FL	60	387	87.8	27.6	75.7	71.7	53.3	6.4
11	m	57	36	54.41	6.3	FL	62	408	88.7	27.2	69.4	82.3	41.9	6.6
12	m	51	15	58.39	35.0	FL	60	283	64.3	21.2	63.3	25.5	7.3	4.3
13	m	51	13	48.71	76.0	FL	64	348	79.5	21.6	68	35.6	22	5.3
14	m	61	0	43.58	7.3	FL	60	353	93.1	24.9	60.2	68.3	33.3	5.7
15	f	65	9	51.14	6.1	FL	60	377	78.7	26.3	63.6	49.2	27.1	6.2
16	m	55	0	45.96	6.1	FL	48	466	62.9	18	51.2	14.6	12.5	7.6
17	m	65	14	43.85	89.9	FL	64	315	88.2	26	58.5	56.2	56.2	4.9
18	m	30	2	38.78	29.0	FL	60	303	86.3	14.5	50	38.3	13.3	4.8
19	m	52	6	48.03	6.3	FL	60	403	90.4	26.1	52.4	85	50	6.7
20	m	44	17	56.79	6.0	FL	58	427	91	22.2	63.2	53.4	13.8	7.2
21	f	27	9	52.25	52.0	FL	46	337	84.2	29.7	64	71.1	28.9	7
22	f	41	5	52.69	16.0	FL	60	369	94.4	21.7	66.2	53.3	28.3	6
23	m	46	3	50.57	32.0	FL	60	386	89.6	31.6	57.4	66.7	58.3	6.4
24	m	57	9	44.67	98.0	FL	60	365	88	16.7	80.3	61.7	48.3	6.1
25	m	65	0	52.57	8.1	FL	34	221	76.5	38.9	65.1	54.5	24.2	6.5
26	m	43	5	58.17	17.1	FL	60	293	76.7	25.3	74.3	53.3	26.7	4.7
27	f	74	7	51.35	6.3	NFL	20	42	34.7	47.6	95	5.3	0	1.7
28	f	45	11	38.15	14.7	NFL	12	46	23	10.9	80	0	0	2
29	f	53	57	47.73	27.9	NFL	60	244	72.8	28.7	62.9	37.9	12.1	4
30	f	45	49	44.75	25.0	NFL	20	77	51	53.2	58.5	0	0	2.8
31	m	41	14	43.13	60.0	NFL	28	97	58.4	35.1	52.9	12	0	2.4
32	f	66	9	49.32	32.3	NFL	60	377	64.2	13.5	58.8	10	21.7	5.3
33	m	49	42	61.04	70.4	NFL	18	78	37.3	38.5	90	46.7	0	3.9
34	f	50	3	48.05	21.4	NFL	22	111	47.2	51.4	50.9	5.6	0	3

35	f	35	11	47.33	25.1	NFL	60	152	55.1	49.3	50.7	3.4	10.3	2
36	m	58	27	46.54	6.3	NFL	52	237	64.9	32.1	36.8	34	0	3.8
37	m	22	29	47.83	59.0	NFL	34	54	60	74.1	80	7.1	0	1.6
38	m	25	25	48.27	64.1	NFL	58	195	53.3	45.6	57.3	18.5	11.1	2.9
39	m	49	23	51.02	45.9	NFL	26	106	45.7	60.4	62.5	15.8	0	3
40	m	62	14	43.02	72.1	NFL	48	251	64.2	54.2	46.3	11.4	0	3.8
41	m	43	18	51.52	85.0	NFL	60	222	72.3	46.8	52.9	14	0	3.1
42	m	34	2	56.91	6.1	NFL	64	297	82	27.3	71.6	32.8	37.5	4.5
43	m	48	7	47.46	47.0	NFL	34	95	45	37.9	77.8	0	0	2.5
44	m	48	23	54.1	28.3	NFL	60	218	82	49.5	61.1	8.9	0	3.3
45	f	43	67	49.26	24.6	NFL	32	115	48.9	40	80.4	33.3	0	3.1
46	m	57	17	44.58	50.9	NFL	60	297	73.5	34.3	50	15	0	4.4
47	m	56	2	46.74	13.1	NFL	48	235	60.7	27.7	60	21.4	0	3.8
48	f	48	86	48.86	21.4	NFL	60	306	81.4	35.9	61.8	33.3	0	4
49	m	65	38	48.81	5.9	NFL	46	208	53.9	54.3	44.2	7.1	0	3.5
50	f	28	16	53.75	55.9	NFL	60	193	61.7	57.5	62.2	3.6	0	3
51	m	61	1	47.17	6.3	NFL	38	190	58.3	28.9	61.8	24.3	10.8	3.8
52	m	30	41	50.3	5.9	NFL	60	145	60.9	69.7	65.3	0	0	2.3

Notes: Part., participant; m, male; f, female; age (yrs), age at t1 in years; dur. (mths), duration of aphasia at t1 in months; severity (T-score), mean profile level of AAT subtest performance (T-scores; reliability-weighted average of T-scores per subtest); t1–t2 (weeks), interval between assessment at t1 and t2 in weeks; type, type of aphasia (FL, fluent (AAT spontaneous speech syntax scale 3 or 4); NFL, non-fluent (syntax scale 0, 1 or 2)); total CLUs, total number of CLUs; total words, total number of open and closed class words; W, percentage words; OCW, percentage open class words; TTR, type-token-ratio; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances.

Appendix 3. Diagnostic Scheme for the Documentation of Individual Patient's Results with Reference to the Normative Value Ranges (from Meffert et al., 2010)

Patient	t1		t2		
Duration of aphasia	CLUs	Words	CLUs	Words	
Parameter (mean; SD; crit. diff.)	Far below avg.	Below avg.	Average	Above avg.	Far above avg.
W (93.3%; 3.5; 6.75)	≤86.2	86.3–89.7	89.8–96.8	≥96.9	–
t1					
t2					
OCW (31.7%; 4.0; 10.71)	≤23.6	23.7–27.6	27.7–35.7	35.8–39.7	≥39.8
t1					
t2					
TTR (82.5%; 6.6; 13.19)	≤69.2	69.3–75.8	75.9–89.1	89.2–95.7	≥95.8
t1					
t2					
COMPL (76.4%; 11.7; 14.76)	≤49.9	53–64.6	64.7–88.1	88.2–99.8	≥99.9
t1					
t2					
CPC (44.3%; 13.8; 15.08)	≤16.6	16.7–30.4	30.5–58.1	58.2–71.9	≥72
t1					
t2					
MLU (6.0; 0.9; 0.99)	≤4.1	4.2–5	5.1–6.9	7–7.8	≥7.9
t1					
t2					

Notes: W, percentage words; OCW, percentage open class words; TTR, type-token-ratio; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances. Critical differences for W, OCW, TTR, COMPL and CPX are given as percentage points, for MLU in words. Value ranges derived from normative data (Meffert et al., 2010): average (mean ± 1 SD); below/above avg., below/above average (mean ± not more than 2 SD), far below/above avg., far below/above average (mean ± at least 2 SD); t1/2, first/second time of assessment.