INT. J. LANG. COMM. DIS., JULY–AUGUST 2008,
VOL. 43, NO. 4, 408–426

informa
healthcare

Research Report

Basic parameters of spontaneous speech as a sensitive method for measuring change during the course of aphasia

Marion Grande†, Katja Hussmann†, Elisabeth Bay†, Swetlana Christoph†, Martina Piefke‡, Klaus Willmes§ and Walter Huber†

†Neurolinguistics, ‡Cognitive Neurology and, §Neuropsychology, Department of Neurology, University Hospital, RWTH Aachen University, Aachen, Germany

(*Received 18 December 2006; accepted 14 September 2007*)

Abstract

Background: Spontaneous speech of aphasic persons is often scored on rating scales assessing aphasic symptoms. Rating scales have the advantage of an easy and fast scoring system, but might lack sensitivity. Quantitative analysis of either aphasic symptoms or basic parameters provides a useful alternative. Basic parameters are essential units of language like word categories or syntactic completeness and can be identified in both impaired and unimpaired adult and child language.
Aims: To examine whether basic parameters of spontaneous speech are more sensitive to change during the course of recovery from aphasia than conventional spontaneous speech rating scales.
Methods & Procedures: Spontaneous speech samples of 28 aphasic participants were analysed using a quantitative computer-assisted method as well as conventional spontaneous speech rating scales before and after 7 weeks of intensive language treatment. The analysis focused on the following basic parameters: percentage words, percentage open class words, syntactic completeness, complexity, and mean length of utterances. The participants were also tested with the Aachen Aphasia Test before and after treatment.
Outcome & Results: Significant change in at least one basic parameter was observed in 20 participants, while only four participants showed significant change in one of the spontaneous speech rating scales.
Conclusions: In comparison with conventional spontaneous speech rating scales, the basic parameters proved to be more sensitive to change. For the time being,

Address correspondence to: Marion Grande, Neurologische Klinik — Neurolinguistik, Universitätsklinikum, D-52057 Aachen, Germany; e-mail: mgrande@ukaachen.de

International Journal of Language & Communication Disorders
ISSN 1368-2822 print/ISSN 1460-6984 online
http://www.informahealthcare.com
DOI: 10.1080/13682820701685991

however, some limitations remain with regard to the specificity of the basic parameters. Thus, additional data are needed to provide further support of the clinical significance of the measured changes.

Keywords: spontaneous speech, aphasia, quantitative analysis, basic parameters.

What this paper adds
Spontaneous speech in aphasia is commonly assessed using rating scales and is usually based on the identification of aphasic symptoms. The purpose of the present study is to apply computer-assisted quantitative analysis to spontaneous speech samples to measure change before and after a period of intensive language treatment. Instead of aphasic symptoms, linguistic basic parameters (e.g. syntactic completeness) were used that also apply to normal language and allow for an easy quantification. The results show that basic parameters proved to be more sensitive to change than the conventional rating scales. Yet, to provide further support of the clinical significance of the measured changes additional data are needed. Nevertheless, the results suggest that computer-assisted analysis of basic spontaneous speech parameters might be a clinically applicable instrument to measure even small changes in spontaneous speech during the course of recovery from aphasia.

Introduction

The ability to participate in and cope with communication in everyday life is strongly influenced by spontaneous speech production. Detailed analysis of spontaneous speech should, therefore, be part of every aphasia examination. Moreover, spontaneous speech should be included when defining treatment goals (Prins and Bastiaanse 2004).

Many clinicians use rating scales from standardized aphasia batteries to assess spontaneous speech production. These rating scales are designed to examine language output on different linguistic and/or communicative levels (Goodglass and Kaplan 1972/1983, Huber *et al.* 1983, Goodglass *et al.* 2001). In spite of being well established in clinical aphasia diagnostics, these rating scales often do not provide sufficient information about an individual's spontaneous speech. The main problem is that they are not sensitive to small changes, which means that changes have to be very substantial to be reflected by the scales. This is often due to poor reliability and/or to rather undifferentiated properties of the rating scales, for example the small number of parameters used to describe the spontaneous speech (for a more detailed discussion of spontaneous speech rating scales, see Prins and Bastiaanse 2004).

Therefore, quantitative methods for spontaneous speech analysis provide an alternative since they allow for a systematic and objective counting of aphasic symptoms and/or basic parameters of speech production (e.g. Shewan 1988; Vermeulen *et al.* 1989; Nicholas and Brookshire 1993; Edwards and Knott 1994; Bastiaanse *et al.* 1996; Berndt *et al.* 2000). Basic parameters are essential units of language and can be identified in every conceivable verbal output (cf. Biniek *et al.* 1991). Thus, they can be assessed in both impaired and unimpaired adult and child language. Basic parameters are defined on different levels (e.g. lexical and syntactic) and differ with respect to the degree of elaborateness (e.g. open/closed class in contrast to a more refined analysis of word classes or even of different types of verbs) depending on the particular research interest.

Saffran *et al.* (1989) (also Berndt *et al.* 2000; Rochon *et al.* 2000), for example, developed quantitative production analysis (QPA) for the detailed analysis of aphasic spontaneous speech. QPA contains several lexical (e.g. total number of words, proportion of closed class words) and syntactic (e.g. degree of sentence elaboration, proportion of sentences containing embedding) basic parameters. The most important aim of the QPA was to develop a method that allows for a detailed description of the special characteristics of agrammatic speech. In a study with 29 non-fluent aphasic speakers, Rochon *et al.* (2000) showed that QPA distinguishes between agrammatic and non-agrammatic aphasic speakers. However, it remained unclear whether the parameters used in the analysis reflected characteristics of agrammatism per se or aphasic deficits in general (also Gordon 2006).

In spite of increasing evidence for the usefulness of quantitative spontaneous speech analysis, the application of these methods is rare in clinical practice. This is primarily due to the long time needed to count and calculate the basic parameters of speech production. To counteract this problem one can revert to computer-assisted methods which allow for quantitative analysis within an acceptable amount of time.

Most of the current computer-assisted methods were developed for the analysis of English spontaneous speech of adults or children (e.g. Crystal *et al.* 1976, Miller and Chapman 1983, MacWhinney 1995). Application to the analysis of aphasic spontaneous speech has been reported, for example by Holland *et al.* (1985), who used a modified version of Systematic Analysis of Language Transcripts (SALT) to analyse the spontaneous speech of a person with global aphasia and could detect changes during the first weeks post-onset. The CLAN programs have also been applied to aphasic spontaneous speech: Murray *et al.* (1998) examined the influence of attention skills on language impairment; and Wright *et al.* (2003) compared different measures of lexical diversity.

With ASPA (Aachener Sprachanalyse; Huber *et al.* 2005) the present authors have introduced a computer-assisted method for the analysis of German spontaneous speech which allows for a detailed assessment of basic parameters within an acceptable amount of time (Hussmann *et al.* 2006). Barthel *et al.* (2006) have proven ASPA to be a valid method with high intra- and interrater reliability. The use of this method requires a transcription of the participant's spontaneous speech, which can be done within the program or by importing an already existing text file. The program analyses the transcript based on several parameters referring to the word and sentence level. The elementary categories are words, interjections and neologisms, open and closed class words, complete, incomplete and elliptic clause-like units (CLUs, see the Methods section for definition), as well as simple and complex CLUs. Each category is indicated both in absolute numbers and in its proportion relative to other categories (e.g. proportion of words in relation to interjections and neologisms). Each category can refer to the whole sample or exclusively to complete, incomplete, or elliptic CLUs (e.g. the proportion of words in complete CLUs). In addition, mean length of utterances (MLU) is indicated (and can also refer to the whole sample or to a special type of CLU), and type–token ratio is calculated separately for open and closed class words.

The paper focuses on the following five basic parameters:

- Percentage words (W), a measure of the lexical content of a speech sample.
- Percentage open class words (OCW), a measure of the semantic content of a speech sample.
- Syntactic completeness (COMPL), i.e. the percentage of syntactically complete CLUs.
- Complexity (CPX), i.e. the percentage of CLUs in compound sentences.
- Mean length of utterances in words (MLU).

The calculation of each parameter is described in more detail in the Methods section (cf. also Hussmann *et al.* 2006).

These parameters were chosen under the assumption that cardinal aphasic symptoms such as word-finding disorders as well as agrammatism and paragrammatism have a clear impact on the distribution of these basic parameters. The authors expect them to reflect different aphasic symptoms and thus to be useful not only for subgroups or types of aphasia. They do not expect to find one parameter to be exclusively sufficient to represent the course of spontaneous speech, but we suppose these parameters to represent sufficiently different aphasic patterns when considered altogether. By the parameter W, for example, different effects of word-finding difficulties should be representable when additional parameters are considered. If aphasic participants respond to their word-finding difficulties by filling the arising pauses by interjections ('the ehm ehm the ehm the … boy'), the value of W in contrast to interjections would decrease. In this case the more severe are the word-finding difficulties, the lower the value of W would be. There are, however, aphasic participants who fill pauses caused by word-finding difficulties by frequently repeating one content word ('the man no man man man no boy'). In this case, the value of W would increase but presumably also OCW and MLU. Therefore, by considering not only one but all parameters, an interpretation should be possible with regard to the severity of the word-finding difficulties and the change over time respectively.

Regarding the parameter OCW and its possible relation to aphasic symptoms, we expect — in addition to word-finding difficulties — particularly agrammatism and paragrammatism to influence this parameter. The more severe agrammatism is, the fewer function words and the more content words will occur. This leads to a high proportion of OCW. Furthermore, agrammatism should evoke rather simple and incomplete phrases. As to the MLU, the effects would depend at least on the participant's response to word-finding difficulties (see above) and on the severity of speech effort. Thus, we would expect the aphasic symptom of agrammatism to be reflected by high OCW, low COMPL, low CPX, and a rather short MLU. Correspondingly, paragrammatism is more likely to cause a low OCW, rather high COMPL, and rather long MLU.

Even though we are aware of the fact that the chosen parameters do not reflect semantic appropriateness or phonological abilities, it is assumed that they altogether do provide useful information about the impact of aphasia on spontaneous speech and consequently may reveal changes in spontaneous speech over time.

The aim of the present study is to apply these basic parameters to the analysis of spontaneous speech in aphasia. By surveying the course of spontaneous speech before and after a period of intensive language treatment, we want to measure change by means of selected basic parameters during aphasia recovery. It is emphasized that we do not intend to differentiate between spontaneous recovery and treatment-induced change; neither do we want to evaluate a specific treatment approach. The comparison of basic parameters with a conventional rating scale is meant to reveal which of the two can demonstrate even subtle change.

The main hypothesis is that basic parameters are more sensitive to change than conventional spontaneous speech rating scales.

Since different patterns of basic parameters are supposed to occur depending on the aphasic symptoms and the coping strategy applied by the individual participant, this study is rather explorative. Thus, hypotheses about particular patterns of change are not straightforward. In general, we assume that improvement of performance on a syntactic level should be reflected in at least one of the syntactic parameters. Improved word retrieval should result in higher W values.

Method

Participants

Twenty-eight aphasic participants were included in the study. Their mean age was 47.4 years (range 22–74 years); their mean duration of aphasia was 18.4 months (range 1–86 months). All had suffered from a cerebro-vascular accident (CVA). Fourteen participants presented with non-fluent aphasia (AAT syntax-scale score 1 or 2), 14 with fluent aphasia (AAT syntax-scale score 3 or 4). At the time of testing, 14 participants (nine fluent, five non-fluent) were in the post-acute stage (1–12 months post-onset) and 14 (five fluent, nine non-fluent) in the chronic stage of aphasia (greater than 12 months post-onset). For descriptive information for the participants, see table 1.

All participants received intensive language treatment on the Aachen Aphasia Ward over a period of 7 weeks, at least 8 hours per week. Treatment goals were defined individually according to an established therapy scheme (Huber *et al.* 1993). This scheme distinguishes three phases of aphasia therapy according to the natural course of aphasia: activation, symptom-specific training, and consolidation (also Huber *et al.* 2006). The participants of the present study were all in the phase of symptom-specific therapy, which aims primarily at relearning degraded linguistic knowledge, reactivating impaired linguistic modalities, and learning compensatory linguistic strategies. There is not a single approach for this phase but a systematic orientation to linguistic units, structures and regularities is essential to all therapeutic attempts. The main focus of each participant's treatment was documented to examine a possible relationship between type of treatment and improvement in basic parameters.

Table 1. Aphasic participants: descriptive information (duration and age at time of pre-test)

Participants	Stage	Total	Male	Female	Duration (months): mean	(range)	Age (years): mean	(range)
Fluent	post-acute	9	8	1	5.6	(1–10)	49.22	(36–68)
	chronic	5	3	2	19.4	(13–36)	50.2	(44–57)
	all fluent	14	11	3	10.5	(1–36)	49.57	(36–68)
Non-fluent	post-acute	5	2	3	6.0	(2–11)	48.2	(34–74)
	chronic	9	3	6	37.4	(18–86)	43.44	(22–58)
	all non-fluent	14	5	9	26.2	(2–86)	45.14	(22–74)
All participants		28	16	12	18.4	(1–86)	47.36	(22–74)

The study was approved by the Research Ethics Committee of the Medical Faculty of the RWTH Aachen University.

Elicitation and transcription of spontaneous speech samples

The AAT was administered before and after 7 weeks of intensive language treatment. All five AAT subtests were taken as an external criterion for improvement since the AAT has been shown to be a psychometrically sound test with strong construct validity, high internal consistency, high test–retest reliability, and high discriminatory power (Huber *et al.* 1983, cf. also Miller *et al.* 2000 for the English version EAAT). The AAT subtests comprise the token test (providing an indication of the overall severity of aphasia), repetition (of sounds, words and sentences with increasing length and complexity), written language (reading, writing and composing from blocks), naming (objects, colours and situations), and comprehension (auditory and reading, words and sentences) (for a more detailed description, see Miller *et al.* 2000).

Spontaneous speech was elicited at the same time using the semi-standardized AAT interview covering four conversational topics (onset of illness, profession, family, and hobbies). The first 60 CLUs were transcribed using ASPA (Huber *et al.* 2005) according to detailed transcription guidelines (Grande *et al.* 2006). To assess interrater reliability, 25% of all transcripts were randomly selected and analysed by a second rater. Interrater reliability was 0.90 for segmentation and 0.95 for classification of CLUs.

If a participant did not produce 60 CLUs, the whole transcript was used (the shortest transcript contained 20 CLUs). This was particularly the case for participants producing non-fluent speech. The mean number of CLUs was 52.7 (range 20–64; as compound sentences were not truncated, some transcripts comprise more than 60 CLUs) for the pre-test and 55.8 (range 25–66) for the post-test.

Transcripts of 60 CLUs were divided half (see below), i.e. the transcripts used for analysis of internal consistency had only 30 CLUs each, a transcript length realizable in clinical daily routine. The transcription and segmentation of 30 CLUs (approximately 2–4 minutes of conversation for fluent and 5–8 minutes for non-fluent speakers) does not take more than 30–45 minutes for an examiner familiar with ASPA.

Quantitative analysis

The CLUs were rated either as complete, incomplete or elliptic. A CLU is defined as a syntactically and/or prosodically marked portion of spontaneous speech referring to a proposition. It is syntactically complete when a verb and all its arguments are given. Only the first elliptic utterance that is produced as response to an open question is counted as an ellipsis (e.g. *How old is your daughter? Ten years*). When marking the CLUs by brackets, the syntactic hierarchy is considered such that subordinated CLUs are identified by the program. Word class (open and closed class words, interjections) is assigned to each item by the program 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 examiner. Open class words are nouns, verbs and adjectives as well as adverbs derived from these word categories.

Closed class words are auxiliaries, determiners, pronouns, conjunctions and genuine adverbs. Interjections are emotional exclamations and onomatopoetic utterances (e.g. *meaouw*). The particles *yes* and *no* are defined as interjections. Words produced with a deviant phonemic structure are counted as neologisms if the target word is not identifiable. Otherwise, the target word is taken.

The program counts the absolute number of items (words, interjections, and neologisms), CLUs, and each of the specified subcategories. Moreover, it calculates several parameters on the word and sentence level.

We focused on the following five basic parameters which were chosen to reveal information about lexical as well as syntactic abilities, particularly about word choice and sentence structure:

- Percentage words (W) in contrast to interjections and neologisms: (words/(words+interjections+neologisms+non-intelligible items)) × 100.
- Percentage open class words (OCW) in contrast to closed class words: (open class words/(open class words+closed class words)) × 100.
- Syntactic completeness (COMPL), i.e. percentage complete 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/number of CLUs).

Qualitative analysis

The transcripts of pre- and post-test were also evaluated by means of the six AAT spontaneous speech rating scales. The ratings scales cover six levels of observation, performance on each level being judged on a six-point scale. Each score on each level is defined by characteristic symptoms and their frequency of occurrence. The features of 'Communicative Behaviour' describe the ability to convey information and to respond to the examiner's questions. 'Articulation and Prosody' covers symptoms of dysarthria and dysprosodia. 'Formulaic Language' includes automatic utterances, verbal stereotypes and echolalia. 'Semantic Structure' comprises semantic paraphasias and neologisms as well as word finding disorders and empty phrases. 'Phonological Structure' is based on the frequency of occurrence of phonemic paraphasias and neologisms. 'Syntactic Structure' covers MLU, syntactic completeness and complexity, absence of function words or inflected forms, and phrase blendings. A detailed description of the respective scores can be found in Miller *et al.* (2000).

Statistical analysis

To test for internal consistency, the transcripts were divided into two subsets, chronologically (first and last half) and randomly. Compound sentences were not truncated, i.e. matrix and embedded clause were always assigned to the same part.

These subparts were analysed separately and formed the basis for estimating reliability and critical differences for significant change. Cronbach's alpha was used as an estimate for the correlation between the two subsamples of one transcript.

Internal consistency was considered sufficient, when it did not fall below 0.80. Critical differences for significant change were calculated according to the following formula:

$$\text{Critical difference} = 1.645 \times \text{SD} \times \sqrt{(2 \times (1 - \text{Cronbach's alpha}))}\,.$$

For each participant significant differences in basic parameters between pre- and post-test were analysed and compared with significant change on the AAT spontaneous speech rating scales. To interpret basic parameter changes in terms of improvement and deterioration, data from a previous study with non-aphasic participants (n=30, 17 female, mean age: 67.7 years; Fiedler 1996) were used (see table A.1 in the appendix).

Two-way analyses of variance (ANOVA) were carried out for each of the five basic parameters to investigate the influence of the three variables type of aphasia (fluent, non-fluent), time post-onset (post-acute, chronic) and severity (moderate, severe; measured by the AAT mean profile level). The repeated measures factor was always time (pre-test, post-test). Altogether, 15 ANOVAs were conducted. The criterion p-value was $p<0.05/15=p<0.003$ (corrected for multiple comparisons).

In order to exclude influences of a particular type of treatment on the improvement in basic parameters, the participants were post-hoc divided into two groups for each type of treatment (participants who had received this particular type of treatment and participants who had not). The two groups were compared by a two-way ANOVA with the repeated measures factor time (pre-test, post-test) and the grouping factor treatment (yes, no). These analyses were conducted only for types of treatment, administered to at least ten and not more than 18 of the 28 participants to ensure that the groups were comparable in size. The types of treatment encompassed in the analysis were written language (administered to 15 participants), syntax (administered to 14 participants), and language comprehension (administered to 12 participants).

Correlational analyses (Pearson; criterion p-value: p=0.05) were carried out among the basic parameters as well as between basic parameters and AAT spontaneous speech rating scales.

The basic parameters were found to be intercorrelated (table 2). W, COMPL, CPX and MLU were positively intercorrelated and negatively correlated with OCW, i.e. participants with a large proportion of words compared with neologisms and interjections showed mostly long as well as complete and complex CLUs and a lower percentage of open class words.

Table 2. Intercorrelations (*r*) between basic parameters of spontaneous speech

Basic parameter	Basic parameter				
	W	OCW	COMPL	MLU	CPX
W	–	−0.60**	0.84**	0.84**	0.73**
OCW	−0.60**	–	−0.64**	−0.76**	−0.50**
COMPL	0.84**	−0.64**	–	0.92**	0.82**
MLU	0.84**	−0.76**	0.92**	–	0.77**
CPX	0.73**	−0.50**	0.82**	0.77**	–

**$p<0.01$.

W, percentage words; OCW, percentage open class words; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances; *r*, correlation coefficient (Pearson).

Table 3. Correlations (*r*) between basic parameters and AAT spontaneous speech rating scales

	AAT rating scale					
Basic parameter	Communicative Behaviour	Articulation and Prosody	Formulaic Language	Semantic Structure	Phonological Structure	Syntactic Structure
W	0.73**	0.42*	0.30	0.11	0.35	0.72**
OCW	−0.42*	−0.46*	−0.17	−0.12	−0.06	−0.79**
COMPL	0.63**	0.44*	0.39*	−0.02	0.26	0.78**
MLU	0.62**	0.45**	0.32	0.04	0.07	0.83**
CPX	0.75**	0.53**	0.45*	0.12	0.29	0.62**

*$p<0.05$; **$p<0.01$.
W, percentage words; OCW, percentage open class words; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances; *r*, correlation coefficient (Pearson).

All five basic parameters were correlated with scores on the AAT spontaneous speech rating scales 'Communicative Abilities', 'Articulation', and 'Syntactic Structure'. Furthermore, COMPL and CPX were positively correlated with 'Automatized Language' (table 3).

Despite the high intercorrelations among the basic parameters, we did not consider it reasonable to regard only one parameter. Instead, the combination of all parameters was expected to provide a comprehensive picture of the participant's language, as was suggested in the Introduction.

As a result of the high correlations between basic parameters and spontaneous speech rating scales, the AAT subtests were included as a measure for general changes in aphasia over time independent of performance in connected speech (and were used for judging the clinical relevance of change in basic parameters).

Results

The results concerning the five basic parameters and the spontaneous speech rating scales are summarized in table 4; significant changes in basic parameters and spontaneous speech rating scales are presented in table 5. Internal consistency was sufficient for all five parameters. Based on the intraclass correlation the critical difference for significant change (at 5% type I error) was determined. Consistency and critical differences are given in table 6. Chronological and random division yielded overall reliability values of identical significance. We derived critical differences only from chronological subsets as this is the more natural division of continuous conversational speech.

Significant change was observed in at least one of the basic parameters in 20 participants. In contrast, four participants showed significant change in at least one of the AAT spontaneous speech rating scales, among them all of the participants showing significant change in at least one of the basic parameters. Eight participants did not show any change at all, neither on the spontaneous speech rating scales nor in the basic parameters. All of the significant changes observed in the spontaneous speech rating scales were increases in the scores achieved and therefore reflected improvement.

Table 4. Results of spontaneous speech using basic parameters and spontaneous speech rating scales at pre- and post-test (n=28)

	Pre-test		Post-test		Number of participants showing a significant change
	Mean	SD	Mean	SD	
Basic parameters					
W (%)	72.0	17.8	74.5	18.7	4
OCW (%)	35.1	13.8	39.8	15.6	4
COMPL (%)	37.6	26.8	41.5	30.3	8
MLU (words)	4.2	1.4	4.5	1.6	4
CPX(%)	17.7	19.1	17.9	17.9	7
Spontaneous speech rating scales[1]					
Communicative Behaviour	2.1	0.8	2.5	0.9	1
Articulation and Prosody	4.0	1.0	4.1	1.0	0
Formulaic Language	3.3	1.1	3.5	1.1	2
Semantic Structure	3.0	0.2	3.1	0.3	0
Phonological Structure	3.2	0.8	3.3	0.8	1
Syntactic Structure	2.2	1.0	2.3	1.1	0

[1]Spontaneous speech rating scales range from zero to five each, where zero was the most severe impairment and five no impairment.
W, percentage words; OCW, percentage open class words; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances.

The definition of improvement in basic parameters is not so straightforward since a significant increase in any one of the five basic parameters may not automatically be rated as improvement. Therefore, an increase was regarded as improvement only for W and COMPL: the higher the proportion of words or complete CLUs, the better. For the other three parameters, judgement is much more dependent on the initial value and their relation to the values of non-aphasic speakers. The classification of changes in these three basic parameters either as improvement or deterioration of performance was therefore based on data from 30 non-aphasic controls from a previous study (Fiedler 1996) (see table A.1 in the appendix). We classified a significant increase in the proportion of OCW, MLU and CPX as improvement if the outcome value was within two standard deviations of the mean of the control group, i.e. 35.6% OCW, MLU of 7.9 words, and CPX of 55.5%.

According to these criteria, 13 of the participants showed improvement in at least one parameter while one participant showed improvement in one parameter (COMPL), but at the same time a possible deterioration of performance in another (CPX). Six showed a possible deterioration of performance in one or more

Table 5. Significant changes from pre- to post-test in basic parameters and spontaneous speech rating scales: number of participants

Qualitative method: AAT spontaneous speech rating scales	Quantitative method: ASPA basic parameters		
	Change	No change	Total
Change	4	0	4
No change	16	8	24
Total	20	8	28

Table 6. Internal consistency coefficients of basic parameters

Parameter	ICC chronological	*p*	ICC random	*p*	Critical difference
W	0.967	<0.001	0.967	<0.001	7.5
OCW	0.811	<0.001	0.947	<0.001	14.0
COMPL	0.957	<0.001	0.965	<0.001	12.9
MLU	0.894	<0.001	0.938	<0.001	1.1
CPX	0.903	<0.001	0.808	<0.001	13.8

ICC, Internal Consistency Coefficient (=Cronbach's alpha); chronological=chronological division of transcripts (first and second half); random=random division of transcripts; W, percentage words; OCW, percentage open class words; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances.
Critical differences for W, OCW, COMPL and CPX are given as percentage points, for MLU in words.

parameters. Table 7 provides an overview of improvement and deterioration in basic parameters, AAT subtests and AAT spontaneous speech rating scales. The issue of improvement and deterioration will be further addressed in the Discussion.

Two-way ANOVAs revealed a significant main effect for type of aphasia (fluent, non-fluent) for all five basic parameters (W: $F(1, 24)=20.35$, $p<0.001$; OCW: $F(1, 24)=30.38$, $p<0.001$; COMPL: $F(1, 24)=56.18$, $p<0.001$; CPX: $F(1, 24)=35.68$, $p<0.001$; MLU: $F(1, 24)=62.23$, $p<0.001$; see table 8 for mean values of each analysis). Fluent participants showed significantly more words (in contrast to interjections and neologisms), fewer open class words, more complete and more complex CLUs, as well as a significantly longer MLU. No significant interaction between time of testing and type of aphasia was found. Consequently, the effect of fluency on basic parameters was not differentially affected by treatment. Furthermore, neither severity (moderate, severe) nor time post-onset (post-acute, chronic) did influence any of the five basic parameters.

To investigate further why some participants did not show change in any basic parameters and whether this was due to the participants' lack of achievement or to the instrument ASPA, the two groups (21 participants showing change and seven participants who did not) were compared with two-sample *t*-tests corrected for multiple comparisons (Bonferroni). No significant differences were found for age, time post-onset (duration in months) or severity (mean AAT profile level).

Furthermore, type of treatment (syntax/written language/comprehension) did not influence changes in basic parameters, since a two-way ANOVA showed no interaction between the two factors time (pre-test, post-test) and type of treatment (treatment administered: yes/no).

Discussion

We expected basic parameters of spontaneous speech production to be more sensitive to small changes in performance than spontaneous speech rating scales. Primarily regarding the mere quantity of change, this was clearly confirmed. While 20 participants showed significant change between pre- and post-test in one or more basic parameters, only four of them had significantly changed in at least one of the rating scales. There are, however, some issues to be discussed when it comes to the quality of the measured changes as the higher sensitivity of the basic parameters may

Table 7. Comparison of changes in basic parameters, AAT subtests and spontaneous speech rating scales for each participant

P	Stage	Severity	Type	Change: basic parameter	Basic parameters: improvement	Basic parameters: deterioration	Change: rating scales	Rating scales: improvement	Change: AAT subtests	AAT subtests: improvement
1	ch	m	fl	*	COMPL		*	PHON	*	REP
2	pa	m	fl	*	COMPL, MLU		*	COM	*	TT, WRI, NAM
3	pa	m	fl	*	MLU				*	TT, NAM
4	pa	s	fl	*	W, COMPL, CPX				*	TT, NAM
5	pa	m	fl	*	CPX				*	TT, REP
6	pa	m	fl	*	COMPL				*	TT, WRI
7	ch	m	nfl	*	W, OCW				*	REP, NAM
8	pa	m	n-fl	*	COMPL	CPX			*	REP
9	pa	m	fl	*		COMPL, CPX			*	NAM
10	pa	s	n-fl	*		OCW			*	TT, REP, WRI, NAM
11	pa	s	n-fl	*		OCW			*	TT, WRI, NAM
12	ch	m	fl	*	CPX		*	AUT		
13	ch	s	n-fl	*		COMPL	*	AUT		
14	pa	s	n-fl	*	W					
15	pa	m	fl	*	CPX					
16	ch	s	fl	*	COMPL					
17	ch	s	n-fl	*	MLU					
18	pa	m	fl	*		CPX				
19	pa	m	n-fl	*		OCW				
20	ch	s	n-fl	*		W, MLU				
21	pa	s	fl						*	TT, REP, NAM
22	ch	m	n-fl						*	NAM
23	ch	s	n-fl							
24	ch	s	n-fl							
25	ch	m	fl							
26	ch	m	fl							
27	ch	s	n-fl							
28	ch	m	n-fl							

pa, Post-acute; ch, chronic, m, moderate, s, severe, fl, fluent, n-fl, non-fluent, *significant change in at least one parameter/subtest/scale; COMPL, syntactic completeness; CPX, complexity; MLU, mean length of utterances; OCW, percentage open class words; W, percentage words; PHO, phonology, AUT, formulaic language, COM, communicative behaviour, TT, token test, REP, repetition, WRI, written language, NAM, naming.

Table 8. Mean values of ANOVAs: influence of the type of aphasia, severity and duration post-onset

	Pre-test		Post-test		Pre-test		Post-test		Pre-test		Post-test	
Parameter	Fluent	Non-fluent	Fluent	Non-fluent	Moderate	Severe	Moderate	Severe	Post-acute	Chronic	Post-acute	Chronic
W	84.2	59.8	85.8	63.2	77.8	64.3	79.0	68.6	74.5	69.6	76.9	72.1
OCW	25.2	45.1	29.0	50.6	32.6	38.6	36.2	44.7	31.3	39.0	37.7	42.0
COMPL	58.2	17.1	66.0	17.0	48.1	23.6	52.5	26.8	43.5	31.8	48.6	34.3
CPX	30.5	4.0	32.6	2.4	25.3	7.0	25.0	8.4	25.2	10.2	23.9	12.6
MLU	5.4	3.0	5.8	3.3	4.7	3.6	5.2	3.7	4.5	4.0	4.9	4.2

CPX, complexity; MLU, mean length of utterances; OCW, percentage open class words; W, percentage words.

possibly adversely affect specificity and the measured changes may not reflect clinically relevant change. These possible limitations will be discussed in the following paragraphs.

One reason for the higher sensitivity of basic parameters might be the fact that each of the spontaneous speech rating scales of the AAT comprises more than one parameter and/or aphasic symptom. For instance, the score on the rating scale 'Syntactic Structure' is defined in terms of MLU, complexity, syntactic completeness and missing or wrong inflection and/or function words. To achieve significant change on this scale, more than one of these features has to improve. Contrarily, separate analysis of each basic parameter revealed isolated improvement, i.e. change of single parameters. Interpretation of changes in basic parameters, however, has been shown to be not so straightforward (see the next section). It might be the case that more data are needed to detect specific patterns of basic parameters for judging the relevance of the changes with regard to the course of aphasia. The clinical applicability of basic parameters will also depend on their explanatory power and level of transparency.

There are some methodological issues to be mentioned. Bird and Franklin (1996) point out the usefulness of quantitative analysis for the comparison of speech samples taken at different times. However, an important aspect in the evaluation of quantitative methods when used for measuring change over time is sufficient reliability since the estimation of the critical difference is based on internal consistency. In this study, internal consistency of each parameter was analysed by way of split-halves method (see above) and found to be high in spite of short transcripts.

However, to confirm the clinical significance of the observed changes, more information about the session-to-session variability would be useful to exclude changes due to participants' motivation, mood or level of fatigue. In further studies, a multiple baseline would be recommended even though the influence of the given factors cannot be completely excluded when not working with a randomized controlled trial which was not possible in this study.

These open questions have to be kept in mind when discussing the results in more detail.

Significant change measured by basic parameters: improvement or deterioration of performance?

Changes observed in basic parameters were classified as improvement or a possible deterioration of performance as described in the results section. Thirteen participants showed changes which were rated as improvement. These were significant increases in W and/or COMPL as well as in MLU, CPX and or OCW which did not exceed the mean of the control group plus two standard deviations. On the other hand, six participants showed deterioration, namely either an increase in OCW above 33% or a decrease in one or more of the other parameters. An increase in the proportion of open class words is always combined with a decrease in the proportion of closed class words. In unimpaired speech of native German speakers, when assessed under AAT conditions, open and closed class words occur in a proportion of about 3:7 (Fiedler 1996). A deviation to one direction or the other may indicate language impairment. Too many open and too few closed class words may reveal morphosyntactic disorders. On the other hand, too few open class words may lead to empty speech.

According to our interpretation, the same holds for MLU and complexity. Both parameters can show too low as well as too high values. An exceedingly high MLU, i.e. atypically long CLUs, may evolve from repetitive phenomena or word-finding difficulties. Too complex sentences mostly occur in connection with paragrammatism while low values of CPX are observed in agrammatism.

When interpreting the changes as improvement or deterioration, interaction among basic parameters on the one hand and between basic parameters and aphasic symptoms on the other hand may play an important role. Additionally, an improvement of not only linguistic but also communicative abilities during recovery from aphasia has to be considered. Thus, an increase in communicative abilities may result in a temporary deterioration of basic parameter values. For instance, in fluent aphasia, a decrease in parameters like COMPL is often observed in connection with increasing inner language control. The participants use fewer paraphasias and verbal stereotypes but at the same time more fragmentary sentences. Nevertheless, the analysis of basic parameters is needed to linguistically quantify the dynamics of recovery and may be helpful in choosing the main focus for a forthcoming treatment interval. For instance, in two aphasic individuals who are both rated with the score of 2 on the AAT-syntax scale ('short, simple sentences, mostly incomplete, almost no function words or inflected forms'), the value of syntactic CPX may indicate whether it is reasonable to start working on complex sentences or to remain on a lower syntactic level for the time being as suggested by reduced syntax therapy (REST; Springer *et al.* 2000).

Altogether, the interpretation of changes measured by basic parameters is possible when using normative data of unimpaired speakers. There are, however, some aspects exceeding the mere linguistic level that cannot be covered by basic parameters. The exclusive use of quantitative basic parameters will not suffice to address the issue of improvement and deterioration in all cases. This will also become clear when considering the clinical relevance.

Clinical relevance of improvement and deterioration

The question remains whether the changes observed in basic parameters are clinically relevant or not. To address this issue, an analysis of individual patterns of performance was carried out with respect to changes in basic parameters, AAT subtests and spontaneous speech rating scales. AAT subtests were included in this analysis because we were interested in general change in the severity of aphasia independent of performance in connected speech. The results given in table 7 display different patterns of change which shall be discussed in the following.

The performance of participants 1 and 2 who show improvement in the AAT subtests as well as in the spontaneous speech rating scales and the basic parameters indicate that in these cases the basic parameters reflect real improvement. There are five participants (numbers 3–7) who exhibit the pattern described in the introduction. Their language abilities improve, as demonstrated by the AAT subtest results, but the AAT spontaneous speech rating scales are not sensitive enough to reflect this improvement. In these cases, it can be assumed that the improvement in basic parameters is clinically relevant. This holds, at least to a certain degree, also for participant 8, who shows improvement in one AAT subtest and in the basic parameter COMPL while at the same time CPX decreases significantly. Here, the increase in COMPL may be a result of more simple instead of embedded sentences.

Participants 9–11 show improvement in up to four AAT subtests, but — according to our criteria — deterioration in one or more basic parameters. Therefore, individual characteristics of each participant's spontaneous speech have to be taken into account in order to get a more comprehensive picture. In the pre-test, participant 9 often gives evasive answers like 'That's a stupid question'. In the post-test, his answers are — presumably due to better lexical access — much more adequate from the communicative point of view. Nevertheless, the semantically adequate answers are more often syntactically simple and incomplete. This example illustrates the problem of interaction between basic parameters and aphasic symptoms as described above. The other two participants show an increase of their already high value of OCW, which can — given the improvement in the AAT subtest 'naming' — also be attributed to better lexical access. A typical answer of participant 10, e.g. when questioned about his hobbies in the pre-test, is 'everything'. In the post-test he answers the same question with 'music Wagner concert'. These cases suggest that better communicative and/or lexico-semantic abilities may involve a deterioration in morpho-syntactic parameters like a decrease of closed class words.

Similar to the cases discussed above, the participants 12 and 13 show a decrease of Formulaic Language and therefore more propositional language which is accompanied by an increase of CPX (in this case a definite improvement) in one participant and a decrease of COMPL (apparently a deterioration of performance!) in the other. The impact of Formulaic Language on basic parameters is an important issue which should be investigated more closely.

The clinical relevance of change in basic parameters is especially difficult to evaluate in those cases where no external evidence, namely improvement in AAT subtests, exists. This is even more the case when deterioration in basic parameters occurs, e.g. in participants 18–20. A closer look at the individual transcripts shows that all three participants show a relatively large deviation from the mean value of the control group in one or two parameters already in the pre-test: Participant 18 exhibits a very complex, paragrammatic sentence structure, participant 19 a very high value of OCW, i.e. severe agrammatism, and participant 20 severe apraxia of speech, resulting in one- and two-word utterances with many neologisms and interjections. Unexpectedly, the deviation of the respective parameter from the mean of the control group is even larger in the post-test. This remains to be analysed thoroughly, presumably not only with quantitative basic parameters as mentioned above.

Participants 21 and 22 show improvement in AAT subtests (token test, repetition and/or naming), but this is reflected neither in the spontaneous speech rating scales nor in basic parameters which indicates a lack of generalization of the trained abilities to spontaneous speech. Participants 23–28 show change in none of the three types of assessment. Here, the absence of change in basic parameters can obviously be attributed to a generally unchanging performance. These six participants are all in the chronic stage, but *t*-tests did not show mean group differences regarding duration post-onset, age or severity. Therefore, the absence of change cannot only be attributed to the fact that these participants were in the chronic stage.

Altogether, the results show some indication of the clinical usefulness of the basic parameters and seem to reflect relevant changes in spontaneous speech at least in some cases. Still, a definite and comprehensive answer to the question of clinical relevance cannot be derived. Consequently, for the time being, it cannot be completely ruled out that basic parameters also reflect non-meaningful change by indeed being sensitive but not specific enough. This has to be examined in more detail in further studies.

Influence of type of aphasia, severity and duration post-onset

The influence of type of aphasia, severity and duration post-onset on the results was tested by means of two-way ANOVAs. Type of aphasia (fluent versus non-fluent) clearly influences the values of all basic parameters. It could be demonstrated that participants with fluent aphasia show higher values for W, MLU, COMPL and CPX and lower values for OCW. Participants with non-fluent aphasia show the reverse pattern. Thus, the basic parameters reliably reflect the AAT grouping criterion for fluent and non-fluent aphasia. This is in line with Gordon (2006) who also reported measures discriminating between fluent and non-fluent speakers. In her study, though, some of the parameters were related to aphasia severity rather than type of aphasia. In the present study, this holds for the syntactic parameters. Participants with fluent and non-fluent aphasia do not, however, differ with respect to the extent of change observed between pre- and post-test. This holds as well for participants with moderate versus severe aphasia and participants in the post-acute versus the chronic stage. This latter finding was unexpected, but the small sample size has to be considered.

Influence of treatment

The type of treatment did apparently not influence change in any basic parameter. This is surprising since a syntactic focus of language therapy could be expected to influence predominantly the syntactic parameters. This absence of interaction between type of treatment and change between pre- and post-test is presumably due to the following methodological problems. On the one hand, type of treatment administered to the individual participant (e.g. semantic, phonological, syntactic) was assessed post-hoc based on the records. Yet, the chosen treatment categories may not have been specific enough for this purpose. Syntactic therapy might train the use of function words or embedded sentences but could also exercise verb–object collocations as used in the reduced syntax therapy (Springer *et al.* 2000). Both would be subsumed under the same type of treatment but should have a quite different impact on the outcome in basic parameters. Furthermore, changes in basic parameters were analysed by means of an ANOVA, i.e. based on mean values so that increases and decreases were levelled out. Clearly, more fine-grained series of single case studies are needed.

Conclusions

In an analysis of spontaneous speech transcripts before and after 7 weeks of intensive language treatment, basic parameters proved to be more sensitive to change than the AAT spontaneous speech rating scales. Interpretation of the measured changes with regard to the quality of change, however, turned out not to be straightforward in all cases. The results indicate that more data are needed to obtain further evidence of specific patterns of change. In ongoing studies more detailed analyses concerning the comparison of aphasic performance with normative data of unimpaired speakers as well as information about the session-to-session variability will be performed. Besides, the question whether, on the one hand, the method of elicitation and, on the other hand, familiarity with the topics chosen affect a participant's performance will be investigated. Given the different patterns of recovery of fluent and non-fluent aphasia, a separate investigation of these types of aphasia should be considered. For this purpose, the analysis should comprise more transcripts.

In spite of these limitations, the authors think they have found some evidence for the assumption that computer-assisted analysis of basic spontaneous speech parameters might be a clinically applicable instrument to measure even small changes in spontaneous speech during the course of recovery from aphasia.

Acknowledgements

This research was supported by the START programme of the Faculty of Medicine, RWTH Aachen University, Aachen. Part of this work was presented at the Clinical Aphasiology Conference 2006 in Ghent, Belgium. The authors would like to thank Delta Systems Aachen for programming the ASPA software. They also thank two anonymous reviewers for helpful comments on an earlier version of the manuscript.

References

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

Bastiaanse, R., Edwards, S. and Kiss, K., 1996, Fluent aphasia in three languages: aspects of spontaneous speech. *Aphasiology*, **10**, 561–575.

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E. and Schwartz, M., 2000, *QPA — Quantitative Production Analysis* (Hove: Psychology Press).

Biniek, R., Huber, W., Willmes, K., Glindemann, R., Brand, H., Fiedler, M. and Annen, C., 1991, Ein Test zur Erfassung von Sprach- und Sprechstörungen in der Akutphase nach Schlaganfällen. *Nervenarzt*, **62**, 108–115.

Bird, H. and Franklin, S., 1996, Cinderella revisited: a comparison of fluent and non-fluent aphasic speech. *Journal of Neurolinguistics*, **9**, 187–206.

Crystal, D., Fletcher, P. and Garman, M., 1976, *The Grammatical Analysis of Language Disability* (London: Edward Arnold).

Edwards, S. and Knott, R., 1994, Assessing spontaneous language abilities of aphasic speakers. *Language Testing*, **11**, 49–64.

Fiedler, M., 1996, Computerunterstützte Analyse der Spontansprache bei chronischen und akuten Aphasien. PhD thesis, RWTH Aachen University.

Goodglass, H. and Kaplan, E., 1972, *The Assessment of Aphasia and Related Disorders* [2nd edn 1983] (Philadelphia, PA: Lea & Febiger).

Goodglass, H., Kaplan, E. and Barresi, B., 2001, *The Assessment of Aphasia and Related Disorders* (Philadelphia, PA: Lipincott, Williams & Wilkins).

Gordon, J. K., 2006, A quantitative production analysis of picture description. *Aphasiology*, **20**, 188–204.

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

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

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

Huber, W., Poeck, K. and Springer, L., 2006, *Klinik und Rehabilitation der Aphasie* (Stuttgart: Thieme).

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

Huber, W., Springer, L. and Willmes, K., 1993, Approaches to aphasia therapy in Aachen. In A. L. Holland and M. M. Forbes (eds), *Aphasia Treatment — World Publishing* (San Diego, CA: Singular), pp. 55–86.

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

MacWhinney, B., 1995, *The CHILDES-Project: Tools for Analyzing Talk* (Hillsdale, NJ: Lawrence Erlbaum).

Miller, J. and Chapman, R., 1983, *Systematic Analysis of Language Transcripts (SALT): A Computer Program Designed to Analyse Free Speech Samples* (Madison: Language Analysis Laboratory, Waisman Center, University of Wisconsin-Madison).

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

Murray, L. L., Holland, A. L. and Beeson, P. M., 1998, Spoken language of individuals with mild fluent aphasia under focused and divided-attention conditions. *Journal of Speech, Language and Hearing Research*, **41**, 213–227.

Nicholas, L. E. and 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. and Bastiaanse, R., 2004, Review. Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, **18**, 1075–1091.

Rochon, E., Saffran, E. M., Berndt, R. S. and Schwartz, M. F., 2000, Quantitative analysis of aphasic sentence production: further development and new data. *Brain and Language*, **72**, 193–218.

Saffran, E. M., Berndt, R. S. and Schwartz, M. F., 1989, The quantitiative analysis of agrammatic production: Procedure and data. *Brain and Language*, **37**, 440–479.

Shewan, C. M., 1988, The Shewan spontaneous language analysis (SSLA) for aphasic adults: description, reliability and validity. *Journal of Communication Disorders*, **21**, 103–138.

Springer, L., Huber, W., Schlenck, K.-J. and Schlenck, C., 2000, Agrammatism: deficit or compensation? Consequences for aphasia therapy. *Neuropsychological Rehabilitation*, **10**, 279–309.

Vermeulen, J., Bastiaanse, R. and Van Wageningen, B., 1989, Spontaneous speech in aphasia: a correlational study. *Brain and Language*, **36**, 252–274.

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

Appendix

Table A.1. Normal range of values from a previous non-aphasic control group (Fiedler, 1996)

Parameter	Mean	Range	Standard Deviation
MLU	5.7	3.4–7.9	1.1
CPX	26.9	0–55.3	14.3
OCW	28.6	23.5–37.6	3.5

MLU, mean length of utterances; CPX, complexity; OCW, open class words.