Published in Clinical Linguistics & Phonetics 31, issue 5, 1-18, 2017,
which should be used for any reference to this work

Inter-individual variability in discourse informativeness in elderly populations

Aurélie Pistono[a,b], Jérémie Pariente[a,c], Catherine Bézy[c], Josette Pastor[a], Thi Mai Tran[d], Antoine Renard[e], Marion Fossard[f], Jean-Luc Nespoulous[b] and Mélanie Jucla[b]

[a]Toulouse NeuroImaging Centre, Toulouse University, Inserm, UPS, Toulouse, France; [b]Octogone-Lordat Interdisciplinary Research Unit (EA 4156), University of Toulouse II-Jean Jaurès, Toulouse, France; [c]Neurology Department, Neuroscience Centre, Toulouse University Hospital, Toulouse, France; [d]"Knowledge, Texts, Language" Research Unit (UMR 8163), Speech Therapy Department, Faculty of Medicine, Lille University, Lille, France; [e]Laboratory for the Exploration of Memory in Neurosciences (LMENS), University of Lausanne (UNIL), Lausanne University Hospital, Lausanne, Switzerland; [f]Institute of Language and Communication Sciences, Neuchâtel University, Neuchâtel, Switzerland

ABSTRACT
An increasing number of studies focus on discourse production in patients with neurodegenerative diseases and underline its clinical usefulness. However, if this is to be used as a clinical tool, one needs to consider how normal discourse varies within cognitively unim-paired elderly populations. In the current study, the aim has been to investigate discourse macrolinguistic variability. For this, 123 participants aged between 55 and 84 were recruited. A cluster analysis of their discourse macrolinguistic features was conducted. Then, cluster characterisation based on socio-demographic and linguistic performance was tested (fluency, naming, syntax and spelling). This method aims to identify various profiles of speaker and informative-ness and then see if inter-individual variability may be related to socio-demographic and/or linguistic aspects. Four clusters of informa-tiveness were found but no socio-demographic features appeared significant. The fourth cluster, defined as 'off topic', had lower performance during linguistic tasks than others and thus the boundary between normality and pathology should be questioned.

KEYWORDS
Cluster analysis; cognition; discourse analysis; inter-individual variability; narrative

Introduction

Discourse within clinical assessment

Increasing numbers of studies are focusing on pathological discourse in a bid to improve current understanding of communication disorders in everyday life (Alzheimer's disease [AD]: Pistono et al. 2016; traumatic brain injury: Coelho et al. 2012; frontotemporal lobar degeneration: Ash et al. 2011). Discourse, being a natural human way of communication, is therefore a crucial concern. Regarding diagnosis for neurodegenerative diseases, discourse is interesting since it allows the assessment of linguistic capacities together with extra-linguistic ones – as neurodegeneration concerns cognition as a whole, a discourse task is considered as essential for clinical assessment (Ska & Duong, 2005).

Thanks to discourse analyses, it has been shown that the discourse of patients suffering from AD presents a lexical–semantic impairment – reduction in semantic content (Ahmed et al. 2013; Forbes-McKay & Venneri 2005) and an increase both in word-finding difficulty and in pausing (Davis & Maclagan, 2009; Forbes-McKay & Venneri, 2005). Patients also present difficulties maintaining a theme while talking (Brandao, 2013; Sharon Ash, Moore, Vesely, & Grossman, 2007). Some results suggest links with memory difficulties associated with speech impairment (Ash et al., 2007; Brandao et al., 2013; Pistono et al., 2016). On the contrary, patients suffering from a frontotemporal lobar degeneration behavioural variant exhibit more difficulties organising their discourse, linked with limited executive resources (Ash et al., 2006), similar to patients suffering from Lewy body spectrum disorders (Ash et al., 2013). In sum, various profiles of discourse may be evidenced according to the neurodegenerative diseases present.

Apart from the fact that discourse assessment is useful in revealing various cognitive impairments and differentiates normal and pathological populations, many authors point out its usefulness for clinical remediation. Indeed, assessing discourse offers perspectives for potential strategies that may compensate for communicative disabilities in neurodegenerative populations, such as strategies for maximising preserved functions or for caregivers to adjust their communication behaviour (Brandao et al., 2013; Carlomagno, Santoro, Menditti, Pandolfi, & Marini, 2005; Dijkstra et al. 2004). These strategies mainly rely on the macrolinguistic features of discourse, which, as defined by Marini et al. (2011), relate to 'pragmatic and discourse-level aspects of language processing, i.e. those recruited in establishing cohesive and conceptual links among contiguous (cohesion and local coherence) or long-distant sentences/utterances (global coherence) and in formulating a mental model or gist of a story or procedure'. Macrolinguistic processing is known to depend upon an extensive cognitive network, and therefore to be more sensitive to ageing (Marini, Boewe, Caltagirone, & Carlomagno, 2005) and neurodegeneration (Glosser & Desser, 1991) than microlinguistic (i.e. phonological, lexical and syntactic processing) tasks. However, if discourse is to be used as a clinical tool, one needs to consider how normal discourse varies within the cognitively unimpaired elderly populations in relation to macrolinguistic processing.

Studying discourse macrolinguistics

Macrolinguistics, as mentioned, relates to the pragmatic aspects of language and encompasses discourse coherence and cohesion. Discourse *coherence* deals with the thematic progression of a discourse and the semantic relationships between propositions regarding the context. It is not simply a discourse property, but an integral part of the concept of discourse (Charolles, 2011). It is crucial to comprehension, as pointed out by Kintsch and van Dijk, a 'language user needs to relate new incoming information to the information he or she already has, either from the text, the context, or from the language user's general knowledge system' (p. 389). The concept of *cohesion*, introduced by Halliday and Hassan (1976), refers to the overt relationships between propositions through cohesive ties. According to these authors, 'cohesion occurs where the interpretation of some elements in the discourse is dependent on that of another' (p. 4).

Currently, two types of analysis are carried out on discourse coherence and cohesion. Some authors consider formal marks (i.e. reference units, Juncos-Rabadán et al., 2005)

and/or errors (Juncos-Rabadán et al., 2005; Marini & Urgesi, 2012) while others propose a scoring of coherence and cohesion, as initially proposed by Glosser and Deser (1991) who scored local and global coherence on a 5 points rating scale [taken up by Kurczek and Duff (2011) or adapted for 4 points rating scale in Wright, Koutsoftas, Capilouto and Fergadiotis (2014)].These scorings reflect how well utterances are linked to the topic and to each other.

Coherence and cohesion are both closely related to discourse informativeness, which corresponds to the ultimate goal of discourse (i.e. conveying a message to an interlocutor). The measure of informativeness, or *information content*, was introduced by Nicholas and Brookshire (1993), based on the fact that 'analyses of the connected speech of adults with aphasia have focused primarily on how their speech conforms to standard language rules and patterns rather than how well it communicates information to listeners' (p. 338). In other words, it was intended to assess information transmission rather than adherence to standard linguistic rules. The authors measured the proportion of *correct information units*, defined as being 'intelligible in context, accurate in relation to the picture(s) or topic, and relevant to and informative about the content of the picture(s) or topic' (p. 348). However, this measure does not adequately show the importance of some units or of any missing information. That is why some authors used units larger than words: Capilouto et al. (2005) carried out a *main event* analysis, which corresponds to key concepts conveying the relation between characters and events. Similarly, Croisile et al. (1996) measured *information units*, which refer to elements of informational content (i.e. actions and entities) that are essential for the narrative. In order to get a broad outlook of discourse macrolinguistic processing and how it may vary, it seems crucial to focus on coherence, cohesion and content information.

Sources of variability in discourse macrolinguistics

As with every cognitive task, inter-individual variability during discourse production occurs according to socio-demographic features. However, contrary to most of existing linguistic tasks, discourse does not lead to single word or sentence production. By definition, discourse is an individual act, which implies considerable variability across subjects in the way this more free-form linguistic task is processed. Apart from the socio-demographic characteristics of the speaker, variation may be related to pragmatic aspects (e.g. topic, interlocutor). For instance, there are many different discourse types (interview, storytelling, narrative, descriptive etc.), and these vary in the amount of cognitive and linguistic load placed on the speaker (Bliss & McCabe 2006).

Concerning variability due to socio-demographic features, most studies observed an ageing effect. Although some authors have reported greater discourse quality in older adults than in younger ones (James, Burke, Austin, & Hulme, 1998), most studies suggest that discourse macrolinguistics decline with age, with fewer correct information units (Capilouto, Wright, & Wagovich, 2005) and greater use of superfluous and irrelevant content units (Cannizzaro & Coelho, 2013; Juncos-Rabadán et al., 2005). Other authors have shown that impairment is even greater in pathological ageing, such as AD (Dijkstra et al., 2004; Kemper, Marquis, & Thompson, 2001; Snowdon et al., 1996). There are fewer results concerning the effect of education on discourse informativeness, although Mulder and Hulstijn (2011) have shown, using four discourse tasks, that participants with a higher

level of education were more successful in conveying their message, talked for longer and made fewer grammatical errors.

As many discourse genres exist, socio-demographic features are not the only sources of variability but relatively few authors have specifically investigated typical discourse variability (for a review, see Obler et al., 1994) in relation to its clinical issues (Smith, Heuerman, Wilson, & Proctor, 2003; Whitworth, Claessen, Leitão, & Webster, 2015). Indeed, patients are usually compared with typical participants but there are few results concerning how normal discourse varies with the demand of the task or how healthy adults structure their discourse. Smith et al. (2003) assessed students' discourse production according to discourse genre, discourse length, participants' sex and oral/written mode. Participants provided more information in personal narratives (in both oral and written mode) than in descriptions, but coherence ratings were consistent across all conditions. They also investigated the gender effect but did not find any significant results. Whitworth et al. (2015) explored discourse production according to discourse genre, discourse topic and participants' age. Coherence was once again stable across genres (recounting, procedure, exposition and narrative), but when the authors looked at different topics within a given genre, they observed subtle differences, with some topics seeming to heighten variability. They also found a decrease in discourse coherence with age. Both studies suggested that discourse macrolinguistics is a clinically useful tool, being a discourse variable stable across conditions. Nevertheless, in each study mentioned, expected sources of variability were explicitly tested, i.e. groups were differentiated by age or gender. These analyses may hide some sources of variability. Moreover, although it is supposed that discourse genres are different regarding the cognitive processes involved, cognitive functions are not taken into account in relation to the success or failure of discourse production.

Objective

In the current study, the aim was to investigate discourse macrolinguistic variability by assessing productions only, using a cluster investigation based on participants' macrolinguistic features (i.e. coherence, cohesion, informativeness). Then, cluster characterisation upon socio-demographic and linguistic performance was tested (fluency, naming, syntax and spelling). This method aims to identify various profiles of speaker and informativeness and then see if inter-individual variability may be related to socio-demographic and/or linguistic aspects. Assessment of macrolinguistic processing together with other linguistic performance items aims to relate informativeness decrease with possible weak performance in other tasks in a typical ageing population. To do so, focus was directed to one genre of narrative within a language assessment battery.

Materials and methods

Participants

A total of 123 healthy participants were recruited, aged between 55 and 84, and equally distributed regarding level of education and sex (Supplementary Table S1). They were all native French speakers, with no history of specific acquired or developmental language

impairment, no uncorrected visual or auditory defect, no general progressive and invalidating disease, no history of neurological or psychiatric disease and no history of alcohol or drug abuse. We controlled for cognition with the Mini-Mental State Examination (MMSE), adjusted for age and education (Kalafat, Hugonot-Diener, & Poitrenaud, 2003). Handedness was evaluated with the Edinburgh test (Oldfield, 1971). All participants gave their informed consent.

Tasks and procedure

Discourse task

Discourse was elicited with a narrative task taken from the GREMOTS language assessment battery (Bézy, Renard & Pariente 2016). The GREMOTS is a computerised language battery dedicated to neurodegenerative diseases at an early stage. This battery evaluates both oral and written language, production and comprehension through 16 tasks. With regard to the narrative task's procedure of the GREMOTS, the experimenter handed three envelopes to the participant, with the following instructions: 'Each envelope contains a story made up of five pictures. Please choose one and tell me the story. I should be able to tell which one you picked up without seeing the pictures'. In fact, the envelopes all contained the same story. This procedure was intended to encourage participants to be precise and to produce as many details as possible. During the task, the experimenter stayed neutral and avoided speaking, in order to have uniform conditions for discourse production.

A sequence of five colour pictures (Supplementary Figure S1) told a story with a clear narrative structure (Labov, 2008): Initial state (pictures 1 and 2), perturbation (picture 3), event itself (picture 4) and final state (picture 5).

All discourse productions were anonymised, digitally recorded and orthographically transcribed.

Other linguistic tasks

For 112 of the 123 participants, data obtained with the full GREMOTS assessment were used. For the current investigation, in addition to the narrative task, seven other tasks in the battery, known to be the most sensitive to pathological ageing, were taken into account – verbal fluency (fruit and verbs), syntactic comprehension, confrontation naming (objects, actions and celebrities) and spelling (Bézy, Renard & Pariente, 2016).

Discourse informativeness analyses

Discourse informativeness was assessed through a chronological grid, corresponding to the presence or absence of different items in each of the four phases of the narrative structure. Those items that were supposed to appear more than once during the story were scored according to the number of expected occurrences. Others were binary variables, characterised by their presence or absence across the whole narrative (Table 1 for examples of each variables under examination and Supplementary Table S2 for whole details).

Most items either improved discourse informativeness (units essential to the narrative structure vs. units characterised as non-mandatory) or impaired it. Other items that reflected the way participants told the story were also examined (i.e. sound illustrations,

Table 1. Discourse informativeness assessment.

Information units improving discourse informativeness			
Respect for the narrative structure		Rating	Examples
Information units	Examples of units under examination		
Presence of the 3 characters	Presence of the father in all 4 phases	4	1: 'The father'; 'Paul'; 'a man with a windsurfing board'
Presence of the place and the main actions in the story	The beach	1	1: 'The story takes place on the beach' 0: No mention of the beach
Presence of correct inferences	Inference of the wind	1	1: 'The wind rises'; 'the palm tree is blown by the wind' 0: No mention of the wind
Non-mandatory elements		Rating	Examples
Information units	Details		
Presence of the secondary actions in the story	7 secondary actions, i.e. mother's and/or son's activities in the 3 phases in which they are present	7	1: 'The mother reads'; 1: 'The mother gives her son a ball'
Elements useful for setting the scene	Mention of the palm tree	1	1: 'The wind blows the palm tree'
Plausible digressions in any of the 4 phases		4	1: 'Last summer'; 'in Nice'
Transitions between the 4 phases		3	1: 'Then'; 'after playing ball'
Information units impairing discourse informativeness			
Referential impairment		Rating	Examples
Modalisations in any of the 4 phases		4	1: 'I'm not sure'; 'apparently'
Ambiguous character references in any of the 4 phases (i.e. each time a reference to a character changes from one phase to another)		9	1: 'Two *friends* go to the beach […] the *father* falls'
Flashbacks		1	1: 'A man is in the water, it's probably the father because at the beginning he had a windsurfing board'
Misinterpretation		Rating	Examples
Mistaken attributions of actions to characters in any of the 4 phases		4	1: 'A man drowns and the father calls for help'
Implausible digressions in any of the 4 phases		4	1: 'The boy says hello to someone in the sea'; 'the mother spends hours on the phone instead of helping her husband'
Other units under examination			
Information units		Rating	Examples
Sound illustration		1	1: 'Aaaah help me!'

For each information unit, one example has been chosen as example. In the examples column, 1 represents correct exemplars of the information unit, while 0 indicates counter examples.

naming characters) to see if they were linked to or independent from discourse informativeness.

To gain more information about discourse production and fluency, the following were also recorded or calculated – the number of words produced, the duration of each narrative and the participants' speech rate (number of words/duration of the narrative).

Inter-rater reliability

Aurélie Pistono (AP) coded all 123 speech samples. However, in order to test for inter-rater reliability, 10 samples were coded by a speech therapist (Catherine Bézy [CB]) and 10 others by a neuro-psycho linguist external to the study. Cohen's kappa coefficient (Cohen, 1968) was calculated for each variable to test the validity of the measurement.

Statistical analyses

Statistical analyses for clustering

Participants' narrative styles were defined using a self-organising Map (Kohonen, 1982). This is an unsupervised artificial neural network designed to arrange multivariate data on a layer of nodes (or neurons) arranged in a two-dimensional grid. It is a self-organised network that detects regularities in the configuration of input data and classifies them accordingly, without residual data. The learning process is competitive: Each neuron in the network has weighted links with the different item values so that when a new training example is presented to the network, it is linked to the closest node, and the network adjusts the weights according to this new input. After training, new patterns are clustered around neurons previously defined by similar training patterns. In the current study, participants were clustered according to informativeness and discourse fluency (discourse length and speech rate), regardless of demographic data or GREMOTS task performance.

The number of expected clusters must be supplied at the start of the analysis. In order to identify the most relevant number of clusters, a first analysis was carried out with six clusters. This did not appear to be appropriate, since two clusters could not be distinguished from others during *post-hoc* analysis. Therefore, to obtain more separate profiles, an analysis was performed on four clusters.

To identify the discourse informativeness and fluency variables on which the clusters were based, they were compared using the non-parametric Kruskal–Wallis or chi-square test for binary ratings. p Values following a significant Kruskal–Wallis test were adjusted for pairwise comparisons (Siegel & Castellan, 1988).

Statistical analyses following clustering

After this initial cluster characterisation, the socio-demographic data and performance on the seven tasks of the GREMOTS battery were entered, to see if they confirmed the clusters. As with the linguistic variables, the non-parametric Kruskal–Wallis or chi-square test was carried out to compare the four clusters. p Values were adjusted for multiple comparisons according to Holm's stepdown procedure (Holm, 1979). Analysis was completed with *post hoc* Mann–Whitney or chi-square tests when necessary.

Concerning performance to the GREMOTS battery, data were also analysed regarding the GREMOTS norms. Although all participants were part of the normative sample, it

allows seeing if participants below the fifth percentile (corresponding to a pathological threshold) were grouped in a same cluster.

Results

Inter-rater reliability

For ordinary ratings, the median kappa coefficient was 0.8 for AP and CB and 0.7 for AP and the neuro-psycho linguist. For binary ratings, the median kappa coefficient was 1 for AP, CB and the neuro-psycho linguist.

Clustering results

Kohonen clustering

The four clusters predicted at the start were highly distinctive. Cluster 1 contained 32 participants, Cluster 2 contained 46, Cluster 3 contained 17 participants and Cluster 4 contained 28 participants. These clusters differed in three of the main categories described in Table 1: Respect of the narrative structure, production of non-mandatory elements, and production of units impairing informativeness. Cluster performance according to these three categories is summarised in Figure 1.

More precisely, five variables defined as essential to the narrative structure were discriminative for cluster specificity: Presence of the father ($H = 28$, $p < 0.001$), presence of the mother ($H = 43$, $p < 0.001$), presence of the son ($H = 61$, $p < 0.001$), presence of correct inferences ($H = 20$, $p < 0.001$), mention of the water activity ($H = 39$, $p < 0.001$). Two of the variables in the non-mandatory elements category were also significant: Secondary actions in the story ($H = 58$, $p < 0.001$) and elements useful for setting the scene ($H = 31$, $p < 0.000$). Finally, two variables impairing informativeness were discriminative: Mistaken attribution of actions to characters ($H = 47$, $p < 0.001$) and ambiguous character references ($H = 35$, $p < 0.001$). None of the variables in the 'other units under examination' category were discriminative (i.e. explicit mention of the medium, sound illustration, use of the first person while talking, naming characters).

For each information category, binary variables, which require a chi-square test, were not significant as every variable had a theoretical rate below 5, preventing the use of this test.

Post-hoc comparisons and cluster specificity for each variable are detailed in Supplementary Table S2.

Discriminative variables following clustering

Discourse fluency: Kruskal–Wallis tests highlighted a significant cluster effect on number of words ($H = 34$, $p < 0.001$), narrative duration ($H = 20$, $p < 0.001$) and speech rate ($H = 9$, $p < 0.05$), as shown in Figure 2.

Socio-demographic data: A main effect of age ($H = 12$, $p < 0.05$) was significant for cluster characterisation. *Post-hoc* analysis evidenced that participants in Cluster 1 were significantly younger than those in Cluster 4. Effects of sex, $\chi^2(3, N = 123) = 1$, $p < 0.74$, and level of education, $\chi^2(6, N = 123) = 6$, $p < 0.46$, were not significant.

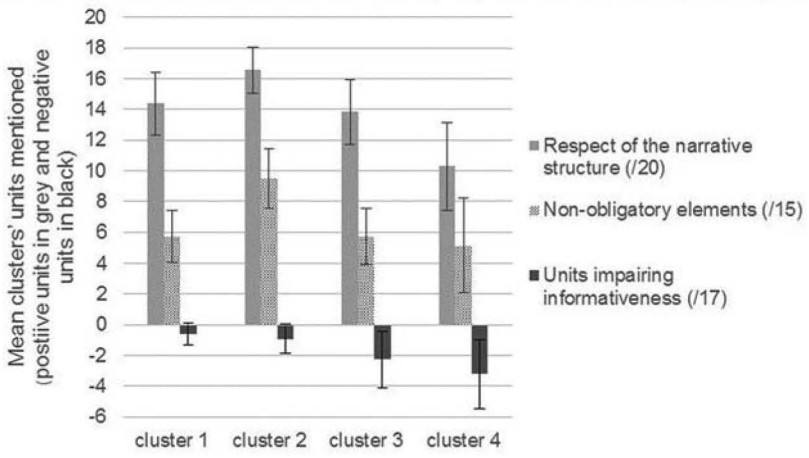

Figure 1. Descriptive analysis of cluster specificity as a function of units improving discourse informativeness (respect of the narrative structure), non-mandatory elements and units impairing informativeness. *Note. Respect of the narrative structure* refers to the presence of the three characters, mention of the water activity and correct inferences (for details, see Supplementary Table S2); *Non-mandatory elements* refer to secondary actions in the story and elements useful for setting the scene (for details, see Supplementary Table S2); *Units impairing informativeness* refer to ambiguous character references and mistaken attributions of actions to characters. Results are shown as negative on the axis to highlight their negative effect on informativeness.

Other language performance: Scores on five of the seven GREMOTS tasks chosen revealed a significant lower performance in Cluster 4, compared with the others (Figure 3): Fruits fluency ($H = 11$, $p = 0.01$), verbs fluency ($H = 20$, $p < 0.001$), syntactic comprehension ($H = 12$, $p < 0.01$), action naming ($H = 13$, $p < 0.01$), object naming ($H = 10$, $p = 0.01$). To check the normality of this sample, the participants' performance was checked against the GREMOTS norms. Participants below the fifth percentile in each task were equally distributed across the four clusters (Figure 3).

Post-hoc analysis and cluster description

Cluster 1 contained informative but concise participants (for an example, see Figure 4). Participants in Cluster 1 talked less in words and duration, but their speech rate did not differ, which means that they were efficient speakers. They hardly generated any errors impairing discourse informativeness (Figure 1). However, they talked less about the son ($z = 6.16$, $p < 0.001$) than those in Cluster 2, less about the mother than Cluster 2 ($z = 3.26$, $p < 0.001$) and Cluster 3 ($z = 3$, $p < 0.01$) and less about the water activity than Cluster 2 ($z = 3.78$, $p < 0.001$) and Cluster 3 ($z = 3.1$, $p < 0.05$). They also produced fewer non-mandatory elements than those in Cluster 2 with fewer secondary actions ($z = 6.54$, $p < 0.00$). In sum, they respected the narrative structure without making any errors but did not extend their narrative to non-crucial elements.

Cluster 2 contained informative and loquacious participants (for an example, see Figure 4). Theirs were the longest discourses in both time and words (Figure 3). They generally produced all the units that were useful to the narrative structure. First, they focused more on the father, the mother and the son than those in the other clusters.

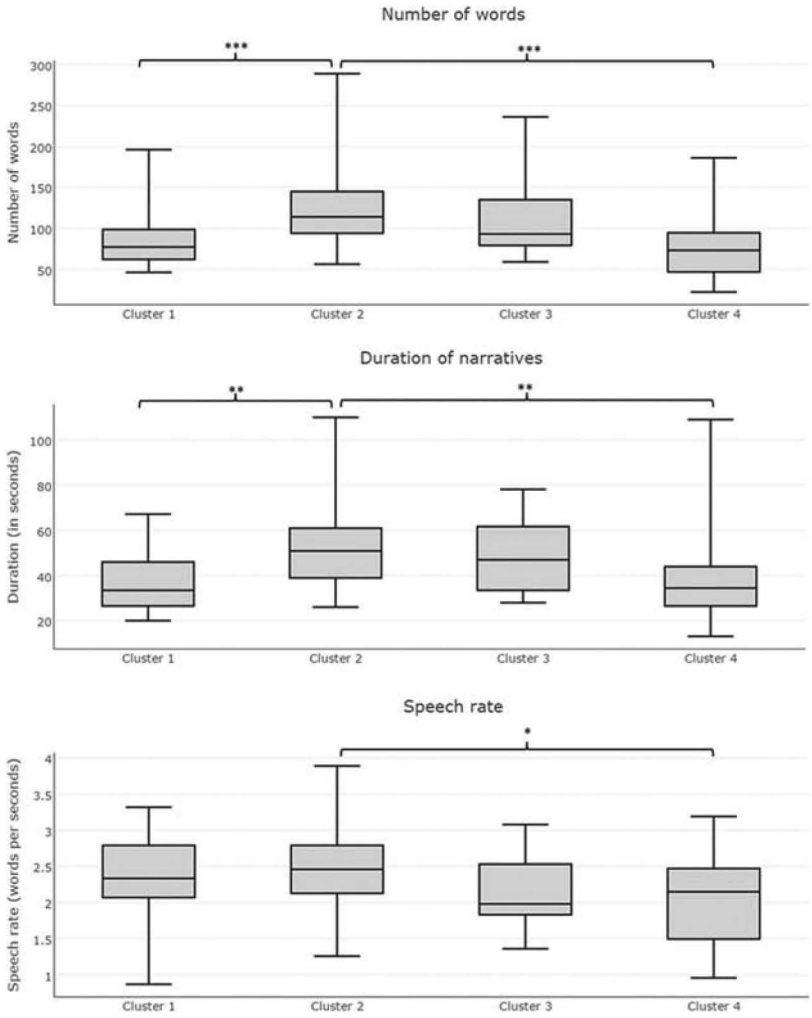

Figure 2. Cluster differences for discourse fluency. *Note. Post-hoc* Mann–Whitney test, corrected for multiple comparisons, for Clusters 1, 2, 3 and 4. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Second, they produced more non-mandatory elements (secondary actions in the story) than participants in Clusters 1, 3 and 4. Participants in this cluster focused on both obligatory and non-obligatory elements (Figure 1).

Cluster 3 contained informative-descriptive participants (for an example, see Figure 4). They rarely significantly differed from Cluster 1 during *post-hoc* analysis but seemed to be less informative. Regarding the narrative structure, they mentioned all the characters, while participants from Cluster 1 focused on the main character (i.e. the father). Participants in Cluster 3 tended to produce more secondary actions in the story and more elements useful for the setting but less inference than participants in Cluster 1. In sum, they mentioned many elements but their discourses were less narrative, using less causal links.

Cluster 4 contained off-topic participants (for an example, see Figure 4). Their speech rate was lower (Figure 2) and they lacked informativeness. They produced less inferences

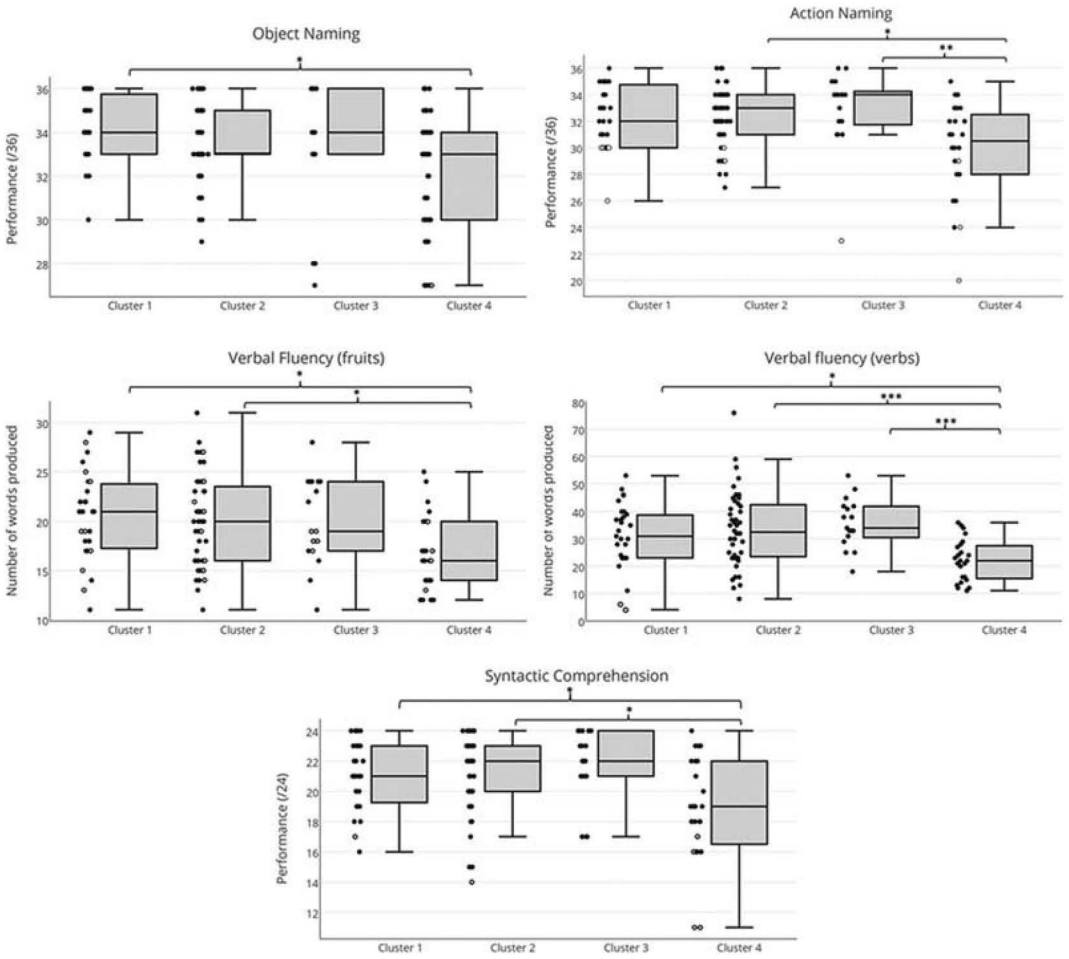

Figure 3. Cluster differences for the GREMOTS variables. *Note. Post-hoc* Mann–Whitney test, corrected for multiple comparisons, for Clusters 1, 2, 3 and 4. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Each participant is represented by a circle; empty circles represent pathological participants.

than Clusters 1 ($z = 4.1$, $p < 0.001$) and 2 ($z = 2.6$, $p < 0.05$), less secondary actions than Cluster 2 ($z = 5.83$, $p < 0.001$) and produced numerous units impairing informativeness. Indeed, they generated many mistaken attributions of actions to characters than those in Clusters 1 ($z = 5.6$, $p < 0.001$), 2 ($z = 5.7$, $p < 0.001$) and 3 ($z = 4.2$, $p < 0.001$) and more ambiguous character references than those in Clusters 1 ($z = 4.2$, $p < 0.001$), 2 ($z = 3.5$, $p < 0.01$) and 3 ($z = 3.7$, $p < 0.01$). Finally, as shown in Figure 3, participants in this cluster differed from the others by their poorer performance on the other language tasks (i.e. verbal fluency, syntactic comprehension, confrontation naming). They were also older than informative–concise participants.

Participant from cluster 1, informative and concise (Man aged 57, level of education 1):
So, today I was on the beach with my parents. I... my father came with me with his beach umbrella. With Mum....Mum brought a ball and toys while my father was in the sea, windsurfing. The sea started to get rough and the wind got up. My father fell into the water. He called for help and Mum called the emergency services.

Participant from cluster 2, informative and loquacious (Woman aged 57, level of education 3):
Well there is a family on holiday. They are at the seaside. And, er, first, we can see the father and his son getting closer to the sea with a beach umbrella and a windsurf board. They settle and then the father goes windsurfing while the little boy stays with his Mum and plays ball. And then they settle...well... the father is still windsurfing, the mother is sunbathing and reading a book. And the boy, after playing, slakes his thirst. At the same time – they do not notice it because they are having fun - but at the same time the winds gets increasingly strong and while the two are sunbathing on the beach, the father is, er, getting into difficulty on his surfboard. He is having real difficulty with the wind and suddenly – bang! He falls into the water. His surfboard turns over and...he is neither swimmer nor an expert, just a holiday-maker and he starts to panic. He shouts for help and his wife and son notice him and start panicking too. And then 4help', the mother calls the emergency services to help her husband.

Participant from cluster 3, informative descriptive (Woman aged 73, level of education 1):
I think it is a family on holiday. Then... the father is with his umbrella and his son. On the next picture, er, no it is not an umbrella... well yes it is. On the next picture the mother is playing with a ball with her son. Well. On the next picture the mother is reading a book. The little boy... we do not really know what he is doing. And the father is on his surfboard. The second... the penultimate picture well in my opinion the father fell and he is struggling, right? And here, well, the mother is scared and is phoning.

Participant from cluster 4, off-topic (Man aged 74, level of education 2):
He is going to the beach. He is going to the beach. He brings his ball. He brings his ball. He is going into the water. And on the beach, it is windy while He is reading. By the way, there are loads of waves. And people are gesticulating to him.

Figure 4. Example of narrative discourse within each cluster.

Discussion

This study has shown that, based on information units, there can be at least four distinct ways of telling the same story in a cognitively unimpaired population aged 55–84. In terms of informativeness, some participants were informative and concise (Cluster 1), while others were informative and loquacious (Cluster 2) or informative-descriptive (Cluster 3). Conversely and surprisingly, a cluster containing 19.5% of all participants was defined as 'off topic' (Cluster 4). Participants from this cluster probably had difficulty with this task – they only mentioned a few of the characters and a few of the actions and did not draw the expected inferences. Moreover, they produced several elements that impaired informativeness, confirming their misinterpretation – ambiguous character references and erroneous attribution of actions to characters. Participants in this cluster were also the only ones to produce a significantly poorer performance in the other linguistic tasks.

These results imply that *healthy controls* are not equally informative when exposed to the same input and the same instructions. Differences were highlighted for most of the variables used, so the fact that the variables reflecting the way participants actually told the story (e.g. use of the first person while talking, giving names to characters) were not discriminative confirms the usefulness of macrolinguistic assessment. The presence of different discourse patterns and inter-individual variability in discourse production have already been found to be influenced by sample, participants' age and discourse genre or topic (Smith et al., 2003; Whitworth et al., 2015). In the current study, although the same constrained task and a larger sample were used, variability still emerged among participants. This variability seems not due to gender or level of education, as clusters did not differ on those variables. However, we found few differences regarding discourse informativeness in ageing, which agrees with previous studies based on discourse macrolinguistic analysis, Cluster 4 being older than Cluster 1 (Capilouto et al., 2005; Juncos-Rabadán et al., 2005; Wright, Koutsoftas, Capilouto, &

Fergadiotis, 2014; Whitworth et al., 2015). Discourse macrolinguistic features seem to decrease with age, even when this effect is assessed afterward and not as an initial hypothesis.

These results were yielded by a clustering investigation, which differs from others in terms of its specific goal. Contrary to most studies, participants were not initially compared on socio-demographic features (e.g. age or level of education), but instead the classification was based on informativeness alone, and looked at why there were similarities and differences between participants. In sum, this method evidenced various ways to process a narrative, exceeding level of education and gender effects. On the contrary, findings on ageing, although significant between only two clusters during *post-hoc* comparisons, may be greater than is suggested in current literature in relation to linguistic performance, as developed below. Indeed, two hypotheses can explain the findings. First, clusters may reflect participants' individual discourse styles, which seem to be more variable among older participants (Obler et al., 1994). Second, they may be linked to language weakness and ageing.

Concerning the *style* hypothesis, participants were exposed to a constrained and unusual discourse situation, to see if there were genuine differences in their way of being informative (e.g. being more or less exhaustive). Instead of focusing directly on participants' formal language abilities, the way they produce informative content was evaluated. This closely corresponded to the definition of cognitive style, which 'concerns individual differences in the processes of cognition, which generally include all processes by which knowledge is acquired: perception, thought, memory, imagery in the "picture-in-the-mind" [...]; it implies the existence of individual differences and preferences in actual modes or manners of gaining, storing, processing, and using information as well' (Ausburn & Ausburn 1977, p. 338).

In other words, the few significant results concerning demographic data suggest that differences may have been due to participants' cognitive style. Authors have already postulated the existence of a cognitive style closely related to discourse coherence. Williams (1985) identified two clusters of speakers corresponding to two cognitive styles: Field-independent participants, who can use decontextualisation and abstraction, and field-dependent participants, who fail to implement these operations and are less coherent. In two informative clusters (1 and 2), the difference lay mainly in the amount of information: One being loquacious and the other being concise, but both used decontextualisation and abstraction with inferences. Cluster 3 may conversely correspond to field-dependent speakers. Participants in Cluster 3, who reported less implicit information, may actually have processed the pictures in a more descriptive manner (i.e. item by item). Comparing various discourse types across the same sample would also be a relevant means of assessing macrostructure consistency across discourse productions, as highlighted in previous studies of variability (e.g. Whitworth et al., 2015) and consistency of the profiles evidenced in the current study.

The finding of 24 'off topic' participants is quite puzzling and raises many concerns, both about normality and about pathology. What is called 'off topic' in the current study refers to participants experiencing referencing and inferencing difficulties and doing misinterpretations. This major issue had already been raised in a clustering study that included patients with AD and healthy matched participants (Duong, Giroux, Tardif, & Ska, 2005). Some mixed clusters pointed to a grey area between normal and pathological

ageing, which needed to be further investigated. To do this, participants' performance in various linguistic tasks was taken into account and it was found that the least informative cluster was also the less proficient one in other linguistic tasks. Such findings highlight the sensitivity of discourse tasks to detecting language weakness since participants in this cluster did not display a clear deficit in linguistic tasks compared with other participants but clearly differed in discourse processing and informativeness. It is not believed that this cluster contained cognitively impaired individuals. First, a check was made for the presence of a global cognitive deficit on the MMSE (it should be noted that there was no cluster effect detected in this task). Second, participants in this cluster did not fall below the pathological threshold of the GREMOTS tasks any more often than the other participants (see Figure 3).

This is why a second hypothesis was suggested, whereby some healthy participants may be *less healthy* than others, even though their performance is still within the norm. Especially regarding language, normality is known to be variable and not unitary. As explained by Nespoulous and Virbel, human verbal behaviour is non-stereotypical and a linguistic disability may be due to many features: 'the linguistic competence of a speaker/listener – pathological as well as normal – can no longer merely be reduced to the control of a certain number of rules – as has long been advocated by many linguistic models' (Nespoulous & Virbel, 2007). In other words, Cluster 4 may be a variant within overall population. Nevertheless, as normality and pathology are not dichotomous but rather represented as a continuum, participants in Cluster 4 may appear at a more distant part of this continuum compared with other participants and represent more than a single variant. More precisely, even if an ageing effect was found (Cluster 4 tended to be older than others and significantly older than Cluster 1), this is not considered to be enough to explain this profile. Indeed, it is now acknowledged that about 30% of healthy people are amyloid positive, which may be a sign of preclinical AD (Mintun et al., 2006; Rowe et al., 2010). At this preclinical stage of AD, subtle cognitive deficits may arise that cannot be captured by the usual standardised tests. Although there is no additional evidence to support this hypothesis, individuals with the poorest communication abilities (assessed here on informativeness) may actually exhibit preclinical AD. The fact that Cluster 4 tended to be older than others is quite coherent as elders are more vulnerable to neurodegenerative diseases. Nor can a cognitive reserve hypothesis to characterise the fourth cluster be discarded. The off-topic participants identified may usually rely on their cognitive reserve, which, as mentioned by Stern (2002), 'should be extended to encompass variation in healthy individuals' performance, particularly when they must perform at their maximum capacity' (p. 448). It can be assumed that the narrative task, which is probably more complex than other linguistic tasks, overloaded participants' maximum capacities. In other words, discourse informativeness assessment may reveal more difficulties than other linguistic tasks, and in a more ecological way. However, although a narrative task may be more sensitive than others, further analyses are required to test its specificity.

To go further, it would be interesting to track this sample during a longitudinal study and apply the clustering to some additional participants, with prodromal AD, to see if they joined the existing clusters (especially, the fourth one) or formed a separate one.

Nonetheless, the above two hypotheses are not mutually exclusive. These results may in part reflect participants' own *style* (especially those in Clusters 1, 2 and 3) or ways of processing the task, but the fourth cluster may represent people at the end of a continuum (Cluster 2 > 1, 3 > 4), the boundary between normality and pathology.

The current findings challenge clinical research, since the frontier between typical and pathological discourse appears vague. Nonetheless, it suggests that focusing on a macro-linguistic analysis of discourse may be a useful clinical tool. It brings different results from other linguistic tasks, highlights the variability across narrative productions and the need to be aware of it. The main findings in the present study show that it is relevant to take account of respect for the narrative structure (mention of the main characters, actions and inferences) and coherence impairment throughout discourse production in this type of analysis. Here, focus was on linguistic tasks which are mainly discriminative for early diagnosis of AD to characterise the clusters further. It can be supposed that using other cognitive tasks together with discourse investigation of patients may be relevant for differential diagnosis, as discourse profiles vary according to neurodegenerative diseases (Ash et al., 2006, 2011, 2013). As mentioned by Roark et al. (2012) in their study on discourse production in AD, it is extremely useful to administer clinical tests that are known to be relevant for studying impaired populations, in order to address specific questions about cognitive processing – which is the reason why the present study focused on this clinical narrative task. It is to be hoped that future studies and/or assessments based on the narrative task will take into account the heterogeneity highlighted in this study. However, it is now crucial to take into account other discourse types, closer to real-life communication, to better comprehend inter-individual variability in relation to cognitive functioning.

Acknowledgments

We thank the participants for giving their time to this study.

Declaration of interest

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

Ahmed, S., Haigh, A. F., de Jager C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*. doi:10.1093/brain/awt269

Ash, S., Avants, B., & Grossman, M. (2011). Non-fluent speech in frontotemporal lobar degeneration, *Journal of Neurolinguistics*, *22*(4), 370–383. doi:10.1016/j.jneuroling.2008.12.001.

Ash, S., Moore, P., Antani, S., McCawley, G., Work, M., & Grossman, M. (2006). Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology*, *66* (9), 1405–1413. doi:10.1212/01.wnl.0000210435.72614.38

Ash, S., Moore, P., Vesely, L., & Grossman, M. (2007). The decline of narrative discourse in Alzheimer's disease. *Brain and Language*, *103*(1–2), 181–182. doi:10.1016/j.bandl.2007.07.105

Ash, S., Xie, S., Gross, R. G., Dreyfuss, M., Boller, A., Camp, E., … Grossman, M. (2013). The organization and anatomy of narrative comprehension and expression in Lewy body spectrum disorders. *Neuropsychology*, *26*(3), 368–384. doi:10.1037/a0027115.

Bézy, C., Renard, A., & Pariente, J. (2016). *GREMOTS Batterie d'évaluation des troubles du langage dans les maladies neurodégénératives*. De Boeck Supérieur: Louvain-la-Neuve.

Bliss, L. S., & Mccabe, A. (2006). Comparison of discourse genres. *Clinical Implications*, *33*, 126–137.

Brandao, L. (2013). Discourse coherence and its relation with cognition in Alzheimer's disease*. *Coherencia del Discurso y su Relación con Cognición en la Enfermedad de Alzheimer Lenisa Brandão*, *7*(1), 99–107. doi:10.5327/Z1982-1247201300010011

Cannizzaro, M. S., & Coelho, C. A. (2013). Analysis of narrative discourse structure as an ecologically relevant measure of executive function in adults. *Journal of Psycholinguistic Research*, *42*(6), 527–549. http://doi.org/10.1007/s10936-012-9231-5

Capilouto, G., Wright, H. H., & Wagovich, S. A. (2005). CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders*, *38*(6), 431–444. doi:10.1016/j.jcomdis.2005.03.005

Carlomagno, S., Santoro, A., Menditti, A., Pandolfi, M., & Marini, A. (2005). Referential communication in Alzheimer's type dementia. *Cortex*, *41*(4), 520–534.

Charolles, M. (2011). Cohreence et cohesion du discours. In K. Holker & C. Marello (Eds.), *Dimensionen der Analyse Texten und Diskursivent - Dimensioni dell'analisi di testi e discoursi* (pp. 153–173). Lit Verlag.

Coelho, C., Lê, K., Mozeiko, J., Krueger, F., & Grafman, J. (2012). Discourse production following injury to the dorsolateral prefrontal cortex. *Neuropsychologia*, *50*(14), 3564–3572. doi:10.1016/j.neuropsychologia.2012.09.005

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4): 213–220. doi:10.1037/h0026256

Davis, B. H., & Maclagan, M. (2009). Examining pauses in Alzheimer's discourse. *American Journal of Alzheimer's Disease and Other Dementias*, *24*(2), 141–154. doi:10.1177/1533317508328138

Dijkstra, K., Bourgeois, M. S., Allen, R. S., & Burgio, L. D. (2004). Conversational coherence: Discourse analysis of older adults with and without dementia. *Journal of Neurolinguistics*, *17*(4), 263–283. doi:10.1016/S0911-6044(03)00048-4

Duong, A., Giroux, F., Tardif, A., & Ska, B. (2005). The heterogeneity of picture-supported narratives in Alzheimer's disease. *Brain and Language*, *93*(2), 173–184. doi:10.1016/j.bandl.2004.10.007

Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences : Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *26*(4), 243–254. doi:10.1007/s10072-005-0467-9

Glosser, G., & Deser, T. (1991). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language*, *40*(1), 67–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2009448

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2): 65–70.

James, L. E., Burke, D. M., Austin, A., & Hulme, E. (1998). Production and perception of "verbosity" in younger and older adults. *Psychology and Aging*, *13*(3), 355–367.

Juncos-Rabadán, O., Pereiro, A. X., & Rodríguez, M. S. (2005). Narrative speech in aging: quantity, information content, and cohesion. *Brain and Language*, *95*(3), 423–434. doi:10.1016/j.bandl.2005.04.001

Kalafat, M., Hugonot-Diener, L., & Poitrenaud, J. (2003). Standardisation et étalonnage français du "Mini Mental State" (MMS) version GRECO. *Rev Neuropsycol.*, *13*(2), 209–236.

Kemper, S., Marquis, J., & Thompson, M. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging*, *16*(4), 600–614.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. http://doi.org/10.1007/BF00337288

Kurczek, J., & Duff, M. C. (2011). Cohesion, coherence, and declarative memory: Discourse patterns in individuals with hippocampal amnesia. *Aphasiology*, *25*(6–7), 700–712. doi:10.1080/02687038.2010.537345.Cohesion

Labov, W. (2008). Oral narratives of personal experience. *Cambridge Encyclopedia of the Language Sciences*, (Maranhão 1984), 546–548. http://doi.org/10.1037/a0022354

Lynna, J., & Ausburn, F. B. (1977). Cognitive styles : Some information and implications for instructional design, *Educational Communication and Technology*, *26*(4), 337–354.

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research*, *34*(5), 439–463. doi:10.1007/s10936-005-6203-z

Marini, A., Galetto, V., Zampieri, E., Vorano, L., Zettin, M., & Carlomagno, S. (2011). Narrative language in traumatic brain injury. *Neuropsychologia*, *49*(10), 2904–2910. doi:10.1016/j.neuropsychologia.2011.06.017

Marini, A., & Urgesi, C. (2012). Please get to the point! A cortical correlate of linguistic informativeness. *Journal of Cognitive Neuroscience*, *24*(11), 2211–22. http://doi.org/10.1162/jocn_a_00283

Mintun, M. A., LaRossa, G. N., Sheline, Y. I., Dence, C. S., Lee, S. Y., Mach, R. H., … Morris, J. C. (2006). [11C]PIB in a nondemented population: Potential antecedent marker of Alzheimer disease. *Neurology*, *67*(3), 446–452. http://doi.org/10.1212/01.wnl.0000228230.26044.a4

Mulder, K., & Hulstijn, J. H. (2011). Linguistic skills of adult native speakers, as a function of age and level of education. *Applied Linguistics*, *32*(May), 475–494. doi:10.1093/applin/amr016

Nespoulous, J-L. & Virbel, J. (2007). From the study of language dysfunction and handicap to a better understanding of linguistic processing in normality. In M. J. Ball & J. S. Damico (Eds.), *Clinical Aphasiology. Future Directions* (pp. 107–124). Hove & New York: Psychology Press, Taylor & Francis Group.

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*(2), 338–350. http://doi.org/10.1044/jshr.3602.338

Obler, L. K., Au, R., Kugler, J., Melvold, J., Tocco, M., & Albert, M. L. (1994). Intersubject variability in adult normal discourse. In R. L. Bloom, L. K. Obler, S. De Santi, & J. S. Ehrlich (Eds.), *Discourse analysis and applications: Studies in adult clinical populations* (pp. 15–27). Hillsdale, NJ: Lawrence Erlbaum.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Pistono, A., Jucla, M., Barbeau, E. J., Saint-Aubert, L., Lemesle, B., Calvet, B., … Pariente, J. (2016). Pauses during autobiographical discourse reflect episodic memory processes in early Alzheimer's disease. *Journal of Alzheimer's Disease*, *50*(3), 687–698. doi:10.3233/JAD-150408

Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K., & Kaye, J. (2012). NIH Public Access, *IEEE Trans Audio Speech Lang Processing*, *7*(3), 263–269. doi:10.1016/j.jalz.2011.03.005.The

Rowe, C. C., Ellis, K. A., Rimajova, M., Bourgeat, P., Pike, K. E., Jones, G., … Villemagne, V. L. (2010). Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. *Neurobiology of Aging, 31*(8),1275–1283. http://doi.org/10.1016/j.neurobiolaging.2010.04.007

Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd Ed.). New York, NY: McGraw-Hill.

Ska, B., & Duong, A. (2005). Communication, discours et démence [Communication Discourse and Dementia]. *Psychologie et Neuropsychiatrie Du Vieillissement*, *3*(2), 125–133.

Smith, R., Heuerman, M., Wilson, B. M., & Proctor, A. (2003). Analysis of normal discourse patterns. *Brain and Cognition*, *53*(2), 368–371. doi:10.1016/S0278-2626(03)00145-3

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA*, *275*(7), 528–532.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society : JINS*, *8*(3), 448–460. doi:10.1017/S1355617702813248

Whitworth, A., Claessen, M., Leitão, S., & Webster, J. (2015). Beyond narrative: Is there an implicit structure to the way in which adults organise their discourse? *Clinical Linguistics & Phonetics*, 1–27. doi:10.3109/02699206.2015.1020450

Williams, J. D. (1985). Coherence and Cognitive Style. *Written Communication*, *2*(4), 473–491. JOUR. http://doi.org/10.1177/0741088385002004008

Wright, H. H., Koutsoftas, A. D., Capilouto, G. J., & Fergadiotis, G. (2014). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *21*(2), 174–196. doi:10.1080/13825585.2013.794894