Role of lexico-syntactic and prosodic cues in spoken comprehension of enumerations in sighted and blind adults

Stéphanie Giraud, Pierre Thérouanne

Role of lexico-syntactic and prosodic cues in spoken comprehension of enumerations in sighted and blind adults

Stéphanie Giraud (1), Pierre Thérouanne (2)

(1) LudoTIC, Nice & Laboratoire de Psychologie Cognitive et Sociale,
Université de Nice – Sophia Antipolis, France
stephanie@ludo-tic.com
(2) Laboratoire de Psychologie Cognitive et Sociale, Université de Nice –
Sophia Antipolis, France
therouan@unice.fr

Abstract

Signaling text organization by different ways may improve comprehension. Two experiments on sighted and blind people studied the benefit provided by signals for spoken language comprehension of expository texts including an enumeration. In addition, these studies tested whether the benefit provided by signals was more important for deep comprehension than for the surface structure of texts. Results showed that comprehension was facilitated when texts were presented with prosodic cues. Moreover, lexico-syntactic signals facilitated comprehension when it required understanding specific semantic relationships between co-enumerated items. However, benefit provided by these signals was restricted to blind participants. Results are discussed in terms of expertise and suggest that signaling should improve access to information for blind people.

Keywords: Comprehension, Spoken language, Text format, Enumeration, Blindness

Note. We thank Léa Pasqualloti and Clément Filippi for their help in designing study and data collection, and the two anonymous reviewers for their helpful comments.

1 Introduction

Information and communication technologies are widely used in everyday life. In one hand, they create a new source of exclusion for people with visual impairment. On the other hand, these technologies potentially fill in their disability by offering access to information with computer interfaces like text-to-speech synthesizers (TTS). Information about the structure of texts conveyed by visual signals of written text should be preserved when texts are oralized by TTS.

1.1 Role of signals in text comprehension

Signaling text organization can be achieved in many ways in written language, with discursive, lexico-syntactic signals (e.g., conjunction), typographic signals (e.g., dash, numbering) and dispositional signals (e.g., horizontal and vertical spacing). According the Textual Architecture Model (Virbel, 1985), signals are realizations of metasentences (e.g., “the first part of the text is…”) describing the elements of the written text itself. In spoken language, intonation, melody, pauses and emphasis may also convey information about text structure.

There is a growing body of evidence that signaling text organization leads to a better comprehension. Lorch and Lorch (1996) showed that headings improve global comprehension of texts. Likewise, Lorch et al. (2001) showed that signals effects on the recall of information result from modifications of text representation are responsible effects of signals. Besides, signals effects varied according expertise level; the benefit provided by signals was more important for novice readers than for expert readers. Schmid and Baccino (2002) showed that formatting text with dispositional signals helped readers to identify perspective shift in narrative texts. Lemarié, Eyrolle and Cellier (2006) showed that discursive and prosodic cues improve the comprehension of restaurant menu: these cues helped them to develop adequate representation of the oralized texts. Finally, Lemarié, Lorch, Eyrolle and Virbel (2008) proposed a model that integrates linguistic and cognitive analyses of signals: SARA (Signal Available Relevant Accessible information). They characterized signals and how signals achieve their effects depending on the availability, relevance and accessibility of information. These authors showed that the magnitude of the effect of a signaling device will increase as its task relevance increases. Signals effects also depend on capacities, knowledge and goals of reader. In addition, discursive signals should be used when the author wants to increase the likelihood that the signaled content will be carefully processed.

However, cognitive processes underlying signals effects need to be investigated. Text comprehension is generally conceived as the successive construction of three levels of representation (Van Dijk, Kintsch, 1983): the *surface structure* consisted of the original words in the text, the *text base* which is the semantic content of the text, and the *situation model* corresponding to the situation described by the text. Different claims have been made about the representational level that would benefit from signaling text organization (see Lemarié et al., 2006). Indeed, Schmid and Baccino (2002) and Maurel et al. (2003) stated that signals leave no trace in the surface structure. Maurel et al. (2003) suggested that signals enable to develop a high level representation through a deep processing, whereas Schmid and Baccino (2002) proposed a fourth level of representation: the *organizational or spatial level*.

1.2 Enumeration

Enumerative structure consists of an introducer, an enumeration listing at least two items, and an optional conclusion (Luc, 2001). The introducer announces the enumeration and can be complete when indicating the number of co-listed items. Furthermore, Luc (2001) distinguished two kinds of enumerations: paradigmatic enumerations listing items that are functionally equivalent, and syntagmatic enumerations exhibiting syntactic or semantic dependence between items.

1.3 The present research

Previous study we conducted showed that reading time of paradigmatic enumerations presented in isolation was shorter when typographic cues (“-“) signaled the co-enumerated items than when lexico-syntactic cues (e.g., “first of all”, “then”, “finally”) were used as signals. Nonetheless, no difference was obtained on comprehension questions. A second experiment did not show any benefit provided by lexico-syntactic cues in spoken language. We made the assumption that texts more complex to process would exhibit a greater benefit from cues. Thus, paradigmatic enumerations were included in larger texts in Experiment 1, and Experiment 2 studied syntagmatic enumerations with semantic dependence between items. Prosodic cues (pauses) and level of comprehension question were also manipulated in Experiment 1.

People with sighted and blind people participated to the two experiments. Previous studies showed that individuals with visual impairment point out better processing of sounds and spoken language than individuals with sight (e.g., Edmonds, Pring, 2006; Röder, Rösler, 2003). Then, we hypothesized that the benefit from cues would be greater for sighted people than for blind people.

2 Experiment 1

2.1 Participants

All participants in the two experiments were French native speakers with no reported hearing difficulties. In Experiment 1, sighted participants were 24 undergraduate and graduate students following a Psychology course, 22 females and 2 males from 18 to 26 years old (median = 22), with normal or corrected sight. Blind participants were 12 members of Valentin Haüy Association and Civil blinds Association. They were 8 females and 4 males aged from 33 to 69 years old (median = 50.5). They were congenitally blind or were blind for at least two years. Their education degree ranged from secondary school to graduate studies.

2.2 Materials

Experimental stimuli consisted of 48 expository texts including a paradigmatic enumeration with 3 or 4 items. Texts were adapted from textbooks for secondary schools and electronic encyclopedias (e.g., Wikipedia) and covered numerous knowledge domains. Enumeration was preceded by an introduction and/or followed by a conclusion. Each enumeration has been created in two versions: an interpretative version and a restricted version (see Figure 1). In the interpretative version, enumeration reconstructs the architectural intentions of author by interpretation; the introducer indicated the number of co-listed items and each item was preceded by a lexico-syntactic cue naming the item category. In the restricted version, no textual cue was given; the introducer was incomplete and lexico-syntactic cues for items were absent. Auditory stimuli were generated with a speech synthesizer (Infovox Desktop, Acapela) using a female voice chosen for its clarity. Prosody was also manipulated. In the version with prosodic cues, the pause ending a sentence lasted 600 ms., the pause ending the introducer lasted 400 ms., and pauses between co-enumerated items lasted 200 ms. In the version with no prosodic cues, all pauses lasted 600 ms.

Text	
Interpretative version	**Restricted version**
Le pingouin est un oiseau noir et blanc de la famille des alcidés, appelé également petit pingouin ou pingouin torda. Les trois caractéristiques différenciant le pingouin du manchot sont énumérées ci-après : la première caractéristique est la zone où il vit, située entre l'Océan Arctique et la Bretagne ; la deuxième caractéristique est sa capacité à voler au dessus de l'eau ; la troisième et dernière caractéristique est sa capacité à plonger en apnée limitée à deux minutes au maximum.	Le pingouin est un oiseau noir et blanc de la famille des alcidés, appelé également petit pingouin ou pingouin torda. Les caractéristiques différenciant le pingouin du manchot sont : la zone où il vit, située entre l'Océan Arctique et la Bretagne ; sa capacité à voler au dessus de l'eau ; sa capacité à plonger en apnée limitée à deux minutes au maximum.
Question	
Surface structure	**Situation model**
Quel mot était dans le texte original ? sa __________ à voler au dessus de l'eau. A. facilité B. compétence C. capacité D. faculté	Où vivent les pingouins ? A. Dans les deux hémisphères B. Uniquement au pôle Nord C. Dans l'hémisphère Nord D. Dans l'hémisphère Sud

Figure 1: Example of Texts and Questions for Experiment 1.

For each text, two questions were elaborated according Daniel and Raney (2007) principles. One correct answer and three alternatives were presented for each question. For the surface structure question, participants had to fill the missing word. For the situation model question, participants had to produce an inference to choose the correct answer. Questions were divided equitably between the introducer and the co-listed items to avoid focusing attention on specific parts of texts. Text version, Prosody, and Question were within-subjects factors.

Six texts without any enumeration were used as fillers, with questions addressing the text base level. Four texts were used for the practice phase, one in interpretative condition, one in restricted condition and two without any enumeration.

2.3 Procedure

Experiment was conducted in laboratory for sighted participants and in a non isolated room outside the laboratory for blind participants. Once the practice phase completed, the 48 experimental and 6 filler trials were presented in a random order. Participants were seated in front of a computer screen and listened texts presented with headphones. Each text was followed by a question. The question and the 4 answers were presented on the computer screen for sighted participants and they were presented with headphones for blind participants[1]. They were instructed to indicate the correct answer by pressing one of four keyboard keys. Then, they pressed the spacebar to start the next trial.

2.4 Results

Correct response percentage was higher for sighted participants (M = 63%) than for blind participants (M = 45%), $F_1(1, 34) = 21.37$, $p < .01$; $F_2(1, 47) = 24.16$, $p < .01$. Correct response percentage was higher with prosodic cues (M = 58%) than with no prosodic cue (M = 52%), although this difference did not reach significance in the subjects' analysis, $F_1(1, 34) = 3.93$, $p = .06$; $F_2(1, 47) = 4.11$, $p < .05$. The four experimental factors - Group, Text version, Prosody, and Question - interacted in the subjects' analysis, $F_1(1, 34) = 6.17$, $p < .05$; $F_2(1, 47) = 3.72$, $p = .06$. No other significant effect was obtained.

		Text version				
		Interpretative		Restricted		
Group	Question	prosodic cues	no cues	prosodic cues	no cues	Mean
Sighted	Surface structure	11007 (65%)	10566 (64%)	11700 (72%)	10741 (58%)	11004 (65%)
	Situation model	17427 (67%)	15633 (56%)	16175 (60%)	16967 (65%)	15114 (62%)
	Mean	14217 (66%)	13100 (60%)	13938 (66%)	13854 (62%)	
Blind	Surface structure	13873 (46%)	15266 (39%)	16639 (50%)	17655 (50%)	15858 (46%)
	Situation model	15567 (44%)	13027 (43%)	20180 (47%)	13541 (39%)	16329 (43%)
	Mean	14720 (45%)	14146 (41%)	18409 (49%)	17098 (45%)	

Table 1: Mean correct response latencies (ms) and percentage of correct responses (in parentheses) as a function of Group, Text version, Prosody and Question. Experiment 1.

Mean correct response latency was shorter in the interpretative version (M = 14045) than in the restricted version (M = 15824), $F_1(1, 33) = 5.67$, $p < .05$. Furthermore, the interaction between the Group and Text version was significant, $F_1(1, 33) = 4.26$, $p < .05$, showing a greater benefit from interpretative version for blind people than for sighted people.

[1] On one hand, procedure should be the same for the two groups. On the other hand, questions were used to test text comprehension itself, and not initial knowledge of participants. Thus, questions were presented in modality which was more convenient for each group, resulting in a variation of procedure between sighted and blind groups. For this reason, any main effect of Group should be interpreted with caution.

3 Experiment 2

3.1 Participants

Sighted participants were 33 undergraduate and graduate students following a Psychology course, 23 females and 10 males from 18 to 35 years old (median = 21.5), with normal or corrected sight. Blind participants were 6 members of Valentin Haüy Association and civil blinds Association. They were 2 females and 4 males from 33 to 66 years old (median = 53). They were congenitally blind or were blind for at least two years. Their education degree ranged from secondary school to graduate studies.

3.2 Materials

Experimental stimuli consisted of 42 syntagmatic enumeration adapted from encyclopedias (e.g., Wikipedia) showing semantic dependence between the 3 or 4 co-enumerated items. Each text has been created in three versions (see Figure 2). The interpretative and restricted version followed the principles described in Experiment 1. In the descriptive version, enumeration described explicitly the textual markers, the introducer also indicated the number of co-listed items, and each item was preceded by a lexico-syntactic cue that did not mention the item category. Text version was a within-subjects factor.

Text		
Interpretative version	Restricted version	Descriptive version
Les quatre manipulations nécessaires à l’observation de l’ADN de l’oignon sont énoncées ci-après : la première manipulation est de couper et broyer les morceaux d’oignon dans un mortier contenant une solution d’extraction ; la deuxième manipulation est de filtrer le broyat obtenu et récupérer le filtrat dans un tube à essai ; la troisième manipulation est d’incliner le tube à essais et verser le long de la paroi le même volume d’alcool à brûler ; la quatrième et dernière manipulation est d’ajouter un colorant afin d’observer l’apparition de filaments.	Les manipulations nécessaires à l’observation de l’ADN de l’oignon sont : de couper et broyer les morceaux d’oignon dans un mortier contenant une solution d’extraction ; de filtrer le broyat obtenu et récupérer le filtrat dans un tube à essai ; d’incliner le tube à essais et verser le long de la paroi le même volume d’alcool à brûler ; d’ajouter un colorant afin d’observer l’apparition de filaments.	Voici une liste de quatre éléments concernant les manipulations nécessaires à l’observation de l’ADN de l’oignon : le premier élément est de couper et broyer les morceaux d’oignon dans un mortier contenant une solution d’extraction ; le deuxième élément est de filtrer le broyat obtenu et récupérer le filtrat dans un tube à essai ; le troisième élément est d’incliner le tube à essais et verser le long de la paroi le même volume d’alcool à brûler ; le quatrième élément est d’ajouter un colorant afin d’observer l’apparition de filaments.
Question (Text base level)		
Lors d’une manipulation visant à observer de l’ADN d’oignon, quand doit-on utiliser un filtre ? A. Après avoir versé l’alcool à brûler. B. Juste après avoir broyé les morceaux d’oignon. C. Juste avant de broyer les morceaux d’oignon. D. Juste avant l’ajout d’un colorant.		

Figure 2: Exemple of Texts and Questions for Experiment 2.

Twelve additional texts were used as fillers, 3 paradigmatic enumerations, 3 syntagmatic enumerations and 6 texts without any enumeration. The practice phase consisted of 2 syntagmatic enumerations and 2 texts without any enumeration.

For each experimental text, one question about the text base level was elaborated according Daniel and Raney (2007) principles. The question dealt with the semantic (i.e., spatial, temporal or causal) relationship between two items. For the fillers, questions never focused on the semantic relationship between items.

3.3 Procedure

The procedure was identical to the procedure of Experiment 1.

3.4 Results

	Text version			
Group	Descriptive	Interpretative	Restricted	Mean
Sighted	12282 (70%)	12388 (70 %)	12343 (66%)	12338 (69%)
Blind	14601 (45 %)	14873 (37%)	15085 (30%)	14853 (37%)

Table 2: Mean correct response latencies (ms) and Percentage of correct responses (in parentheses) as a function of Group and Text version. Experiment 2.

Correct response percentage was higher for sighted participants than for blind participants, $F_1(1,37) = 38.73$, $p<.001$; $F_2(1,41) = 61.03$, $p<.001$. Main effect of Text version was significant in subject's analysis, $F_1(2,74) = 3.52$, $p<.05$; $F_2 < 1$. Planned comparisons showed more correct responses in the descriptive and interpretative conditions than in the restricted condition, although this difference was not significant in the items' analysis, $F_1(1,37) = 5.22$, $p<.05$; $F_2(1,41) = 1.54$, $p>.10$, whereas the difference between the descriptive condition and the interpretative condition was not significant, $F_1(1,37) = 1.50$, $p>.10$; $F_2 < 1$. Group and Text version did not interacted, $F_1(2,74) = 1.32$, $p > .10$; $F_2 < 1$.

Mean correct response latency was shorter for sighted participants than for blind participants, although this difference did not reach significance, $F_1(1, 37) = 38.34$, $p = .08$. No other effect was significant.

4 Discussion

Experiment 1 did not show any benefit provided by lexico-syntactic cues on comprehension of paradigmatic enumeration. However, Experiment 2 showed such a benefit on syntagmatic enumerations, correct responses being more frequent in interpretative and descriptive versions of texts than in restricted version. This pattern of result confirms our first hypothesis and strongly suggests that lexico-syntactic cues facilitate comprehension when establishing the specific semantic relationships between co-enumerated items is necessary. Experiment 1 also showed that responses were facilitated when texts were presented with prosodic cues, despite the fact that these cues were somewhat subtle variation of pauses duration. However, Experiment 1 failed to show that facilitation provided by signals is greater for situation model than for surface structure.

As blind people rely mainly on spoken language, we predicted that they would less beneficiate from cues than participants with sight. On the contrary, the sighted group was more accurate than the blind group in the two experiments. Moreover, the benefit provided by cues was higher on response latencies for blind participants than for participants with sight in Experiment 1. Although they could be considered as more expert in spoken language activities, blind participants were also older with educational degree largely lesser than participants with sight. Indeed, we failed to match blind and sighted people on main characteristics relevant to this study. We hypothesize that blind participants had fewer background knowledge dealing with the content of the presented texts, making signaling more useful to them (see Lemarié et al, 2008).

Two research directions will be followed. First, given the relations between age, access to knowledge and blindness, it seems that signaling is helpful to improve accessibility of information presented by the mean of spoken language, especially with TTS synthesizers. Second, for a more theoretical purpose, any comparison between blind and sighted participants will require matching on background knowledge and sample size despite the fact that it is difficult to achieve.

References

DANIEL F., RANEY, G. E. (2007). Capturing the effect of a title on multiple levels of comprehension. *Behavior Research Methods* 39, 892-900.

EDMONDS C. J., PRING L. (2006). Generating inferences from written and spoken language: A comparison of children with visual impairment and children with sight. *British Journal of Developmental Psychology* 24, 337-351.

LEMARIÉ J, EYROLLE H., CELLIER J.-M. (2006). Visual signals in text comprehension: How to restore them when oralizing a text via a speech synthesis? *Computers in Human Behavior* 22, 1096-1115.

LEMARIÉ J., LORCH R. F., EYROLLE H., VIRBEL J. (2008). SARA: A text-based and reader-based theory of text signaling. *Educational Psychologist* 43, 1-23.

LORCH R. F., LORCH E. P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology* 21, 261-278.

LORCH R. F., LORCH E. P., RITCHEY K., MCGOVERN L., COLEMAN D. (2001). Effects of headings on text summarization. *Contemporary Educational Psychology* 26, 171-191.

LUC C. (2001). Une typologie des énumérations basée sur les structures rhétoriques et architecturales du texte. Proceedings of *Conférence Traitement Automatique du Langage Naturel (TALN'2001).*

MAUREL F., LEMARIE J., VIGOUROUX N. (2003). Evaluation cognitive d'une représentation du texte pour sa présentation multimodale. Proceedings of *Conférence Internationale sur le Document Electronique (CIDE 6).*

RÖDER B., RÖSLER F. (2003). Memory for environmental sounds in sighted, congenitally blind and late blind adults: Evidence for cross-modal compensation. *International Journal of Psychophysiology* 50, 27-39.

SCHMID S., BACCINO T. (2002). Perspective shift and text format: An eye-tracking study. *Current Psychology Letters: Behaviour, Brain & Cognition* 9, 73-87.

VAN DIJK T. A., KINTSCH W. (1983). *Strategies of discourse comprehension*. New York: Academic.

VIRBEL J. (1985). Langage et métalangage dans le texte du point de vue de l'édition en informatique textuelle. *Cahiers de Grammaire* 10, 1-72.