

WIP: Integrating Text and Graphics Design for Adaptive Information Presentation

Wolfgang Wahlster, Elisabeth André, Wolfgang Finkler, Winfried Graf,
Hans-Jürgen Profitlich, Thomas Rist, Anne Schauder
German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
D-6600 Saarbrücken 11
Germany
Email: *last_name*@DFKI.UNI-SB.DE

When explaining how to use a technical device humans will often utilize a combination of language and graphics. It is a rare instruction manual that does not contain illustrations. Multimodal presentation systems combining natural language and graphics take advantage of both the individual strength of each communication mode and the fact that both modes can be employed in parallel. It is an important goal of this research not simply to merge the verbalization results of a natural language generator and the visualization results of a knowledge-based graphics design component, but to carefully coordinate natural language and graphics in such a way that they generate a multiplicative improvement in communication capabilities. Allowing all of the modalities to refer to and depend upon each other is a key to the richness of multimodal communication.

In the WIP system that plans and coordinates multimodal presentations in which all material is generated by the system, we have integrated multiple AI components such as planning, knowledge representation, natural language generation, and graphics generation. The current prototype of WIP generates multimodal explanations and instructions for assembling, using, maintaining or repairing physical devices. As we try to substantiate our results with cross-language and cross-application evidence, WIP is currently able to generate simple German or English explanations for using an espresso-machine or assembling a lawn-mower.

In WIP we combined and extended only formalisms that have reached a certain level of maturity: in particular, terminological logics, RST-based planning, constraint processing techniques, and tree adjoining grammars with feature unification.

One of the important insights we gained from building the WIP system is that it is actually possible to extend and adapt many of the fundamental concepts developed to date in AI and computational linguistics for the generation of natural language in such a way that they become useful for the generation of graphics and text-picture combinations as well. This means that an interesting methodological transfer from the area of natural language processing to a much broader computational model of multimodal communication seems possible. In particular, semantic and pragmatic concepts like coherence, focus, communicative act, discourse model, reference, implicature, anaphora, or scope ambiguity take an extended meaning in the context of text-picture combinations.

A basic principle underlying the WIP model is that the various constituents of a multimodal presentation should be generated from a common representation of what is to be conveyed. This raises the question of how to decompose a given communica-

tive goal into subgoals to be realized by the mode-specific generators, so that they complement each other. To address this problem, we explored computational models of the cognitive decision processes coping with questions such as what should go into text, what should go into graphics, and which kinds of links between the verbal and non-verbal fragments are necessary.

The task of the knowledge-based presentation system WIP is the context-sensitive generation of a variety of multimodal documents from an input including a presentation goal. The presentation goal is a formal representation of the communicative intent specified by the back-end application system. WIP is a highly adaptive interface, since all of its output is generated on the fly and customized for the intended target audience and situation. The quest for adaptation is based on the fact that is impossible to anticipate the needs and requirements of each potential user in an infinite number of presentation situations. We view the design of multimodal presentations including text and graphics design as a subarea of general communication design. We approximate the fact that communication is always situated by introducing generation parameters (see Fig. 1) in our model. The current system includes a choice between user stereotypes (e.g. novice, expert), target languages, layout formats (e.g. hardcopy of instruction manual, screen display), and output modes (incremental output vs. complete output only). The set of generation parameters is used to specify design constraints that must be satisfied by the final presentation.

The technical knowledge to be presented by WIP is encoded in RAT (Representation of Actions in Terminological Logics). In addition to this propositional representation, WIP has access to an analogical representation of the geometry of the machine in the form of a wireframe model (see Fig. 1). Note that the incrementality

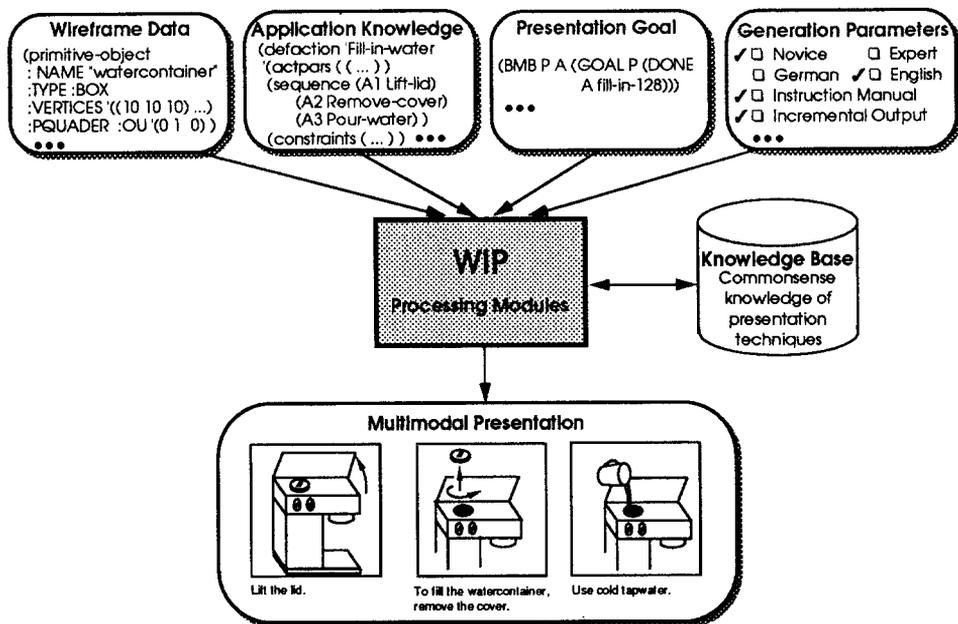


Fig. 1. A Functional View of the WIP system

mentioned above as one of the options for the generation of multimodal output, characterizes a likely application scenario for systems like WIP, since the intended use includes intelligent control panels and active help systems, where the timeliness and fluency of output is critical, e.g. when generating a warning. In such a situation, the presentation system must be able to start with an incremental output although it has not yet received all the information to be conveyed from the back-end system.

References: Selected Papers from the WIP Project

- [André and Rist 90a] André, E. and Rist, T. 1990. Towards a Plan-Based Synthesis of Illustrated Documents. In *Proceedings of the 9th ECAI* (25–30).
- [André and Rist 90b] André, E. and Rist, T. 1990. Synthesizing Illustrated Documents: A Plan-Based Approach. In *Proceedings of InfoJapan '90* Vol. 2 (163–170).
- [André and Rist 92] André, E. and Rist, T. 1992. The Design of Illustrated Documents as a Planning Task. German Research Center for Artificial Intelligence, DFKI Research Report.
- [André et al. 92] André, E., Finkler, W., Graf, W., Rist, T., Schauder, A., and Wahlster, W. 1992. WIP: The Automatic Synthesis of Multimodal Presentations. German Research Center for Artificial Intelligence, DFKI Research Report.
- [Finkler and Schauder 92] Finkler, W. and Schauder, A. 1992. Effects of Incremental Output on Incremental Natural Language Generation. German Research Center for Artificial Intelligence, DFKI Research.
- [Graf 92] Graf, W. 1992. Constraint-Based Graphical Layout of Multimodal Presentations. German Research Center for Artificial Intelligence, DFKI Research Report.
- [Harbusch 90] Harbusch, K. 1990. Constraining Tree Adjoining Grammars by Unification. in *Proceedings of the 13th COLING* (167–172).
- [Harbusch et al. 91] Harbusch, K., Finkler, W. and Schauder, A. 1991. Incremental Syntax Generation with Tree Adjoining Grammars. In Brauer, W. and Hernández, D. (eds.) *Proceedings of the 4th International GI Conference on Knowledge-based Systems*. Berlin: Springer-Verlag (363–374).
- [Rist and André 92] Rist, T. and André, E. 1992. From Presentation Tasks to Pictures: Towards an Approach to Automatic Graphics Design. German Research Center for Artificial Intelligence, DFKI Research Report.
- [Schauder 92] Schauder, A. 1992. Incremental Syntactic Generation of Natural Language with Tree Adjoining Grammars. German Research Center for Artificial Intelligence, DFKI Research Report.
- [Wahlster et al. 89] Wahlster, W., André, E., Hecking, M., and Rist, T. 1989. WIP: Knowledge-based Presentation of Information. German Research Center for Artificial Intelligence, DFKI Research Report.
- [Wahlster et al. 91] Wahlster, W., André, E., Graf, W., and Rist, T. 1991. Designing Illustrated Texts: How Language Production Is Influenced by Graphics Generation. In *Proceedings of the 5th Conference of the European Chapter of the ACL* (8–14).
- [Wahlster et al. 92a] Wahlster, W., André, E., Bandyopadhyay, S., Graf, W., and Rist, T. 1992. WIP: The Coordinated Generation of Multimodal Presentations from a Common Representation. In Ortony, A., Slack, J., and Stock, O. (eds.), *Computational Theories of Communication and their Applications*. Berlin: Springer-Verlag.
- [Wahlster et al. 92b] Wahlster, W., André, E., Finkler, W., Profitlich, H.-J., and Rist, T. 1992. Plan-based Integration of Natural Language and Graphics Generation. German Research Center for Artificial Intelligence, DFKI Research Report.
- [Wazinski 92] Wazinski, P. 1992. Generating Spatial Descriptions for Cross-modal References. In *Proceedings of the 3rd ACL Conference on Applied Natural Language Processing*.