Sonderforschungsbereich 314

Künstliche Intelligenz - Wissensbasierte Systeme

KI-Labor am Lehrstuhl für Informatik IV

Leitung: Prof. Dr. W. Wahlster

Universität des Saarlandes FB 14 Informatik IV Postfach 151150 D-66041 Saarbrücken Fed. Rep. of Germany

Tel. 0681 / 302-2363



Bericht Nr. 92

VITRA GUIDE:

Utilisation du Langage Naturel et de Représentations Graphiques pour la Description d'Itinéraires.

Gerd Herzog, Wolfgang Maaß, Peter Wazinski

Januar 1993

ISSN 0944-7814 **92**

VITRA GUIDE:

Utilisation du Langage Naturel et de Représentations Graphiques pour la Description d'Itinéraires.*

Gerd Herzog, Wolfgang Maaß, Peter Wazinski SFB 314, Projet VITRA Universität des Saarlandes D-6600 Saarbrücken 11 R.F.A. vitra@cs.uni-sb.de

janvier 1993

Résumé

En tant que projet de recherche de base, nous sommes concernés dans VITRA (VIsual TRAnslator) par la connection entre langage et perception visuelle. Visant à l'élaboration d'un système de représentation des connaissances qui permette un accès en langage naturel aux données visuelles, nous sommes plus particulièrement intéressés par la génération automatique de descriptions d'itinéraires. Afin de générer des descriptions d'itinéraires adequates aux besoins de l'utilisateur et d'applications spécifiques, nous utilisons les modes de présentation suivants : descriptions en langage naturel, cartes 2D et graphiques 3D. Nous présentons dans cet article¹ ces différents modes de présentation et montrons comment ils peuvent être combinés pour une description multi-modale d'itinéraires.

Abstract

As a basic research project, in VITRA we are concerned with the connection between language and visual perception. Aiming at knowledge-based systems that provide a natural language access for visual data, we especially deal with the automatic generation of route descriptions. In order to produce route descriptions taylored to user and application specific needs we use the following presentation modes: natural language descriptions, 2-dimensional maps and 3-dimensional graphics. In this paper we discuss these presentation modes and show how they can be combined for a multimodal route description.

^{*}Dans : Colloque Interdisciplinaire du Comité National "Images et Langages: Multimodalité et Modélisation Cognitive" Paris 1993

^ĨCe travail est sponsorisé par le Sonderforschungsbereich 314 der Deutschen Forschungsgemeinschaft, "Künstliche Intelligenz und wissensbasierte Systeme", projet N2: VITRA.

1 Introduction

Nous analysons dans le projet VITRA (VIsual TRAnslator) la jonction de la compréhension d'images et du traitement automatique du langage naturel (LN). A cet effet nous avons effectué des études experimentales pour le développement de systèmes à bases de connaissances (cf. [Wahlster 87]). En plus de la description en langage naturel de séquences d'images (cf. [Herzog et al. 89]), où il s'agit de l'intégration des systèmes de traitement du langage naturel aux systèmes de compréhension d'images, nous nous intéressons à l'élaboration automatique en LN de description d'itinéraires.

Plusieurs études linguistiques et psycho-linguistiques traitent la question de la description d'itinéraires (cf. par ex. [Wunderlich & Reinelt 82], [Klein 83] et [Rauh & Schweitzer 87]). Beaucoup de travaux en intelligence artificielle abordent aussi ce problème (cf. par ex. [Habel 87], [Davis 89], [Hoeppner et al. 90] et [Ligozat 92]).

L'approche suivie jusqu'à présent dans VITRA se limitait à une interaction en LN entre le système et l'utilisateur, consistant à répondre aux questions de ce dernier concernant les relations spatiales et les trajectoires pour une séquence d'images codifiées dans un modèle géometrique bi-dimensionnel (cf. [Müller 89]). Nous traitons dans le nouveau système VITRA GUIDE la combinaison des représentations graphiques et du langage naturel. Nous utilisons à cette fin un modèle géometrique tri-dimensionnel d'un fragment du monde qui permet une description incrementale des itinéraires. La génération du texte de ces descriptions est réalisée en prenant en compte le parcours correspondant, qui est suivi graduellement, et du contexte visuel.

Dans le travail présenté ici, la transmission d'informations spatiales occupe le premier plan. C'est un élément central dans la description des itinéraires. Nous montrerons dans quelle mesure on peut combiner les représentations graphiques aux aspects linguistiques d'une configuration spatiale.

2 Description en langage naturel des relations spatiales

La transmission d'informations spatiales lors de la description d'itinéraires sert à l'identification et à la caractérisation de morceaux de chemins, de marqueurs spatiaux et à l'indication de changements de direction²:

- (1) Gehen Sie an der nächsten Kreuzung rechts. "Tournez à droite au prochain croisement."
- (2) Bleiben Sie auf der Straße bis Sie *links von* ihnen ein Gebäude mit einem direkt davor liegenden Parkplatz sehen.
 - "Restez sur la rue jusqu'à ce que vous voyiez sur votre gauche un bâtiment avec un parking juste devant."
- (3) Dahinter wählen Sie den links hinauf führenden Weg. "Après ce bâtiment, prenez à gauche le chemin montant."
- (4) Gehen Sie an der Rasenfläche entlang, unter der Brücke hindurch bis zur Bushaltestelle. "Allez longez le gazon en passant sous le pont jusqu'à l'arrêt du bus."
- (1) décrit un changement de direction, alors que dans (2) un marqueur spatial est introduit à l'aide de prépositions d'orientation. Nous utilisons dans (3) une expression complexe de référence pour le choix d'un itinéraire. A cet effet, nous combinons deux prépositions d'orientation (à gauche, en haut). (4) décrit deux morceaux de trajectoires par le biais de deux prepositions faisant référence à des chemins (le long et sous).

 $^{^2}$ Dans les exemples suivants, nous essayerons de donner une traduction en français aussi fidèle que possible.

Dans VITRA GUIDE la connaissance spatiale est représentée implicitement sous la forme d'un modèle géometrique. La génération d'expressions de référence spatiale présuppose la transformation de la représentation géometrique en une description explicite de la configuration spatiale.

L'analyse référentielle et sémantique des prépositions spatiales conduit au concept de relation spatiale dont la sémantique est indépendante du langage étudié.

Les relations spatiales sont définies par la spécification de conditions sur les données spatiales servant à la configuration d'objets, comme, par exemple, la distance entre deux objets, la position relative par rapport à une orientation; c'est-à-dire on caractérise une certaine classe de configuration d'objets.

A cet effet, il faut tenir compte de la distinction entre l'applicabilité et la non-applicabilité d'une expression. Mais en général, cette distinction ne suffit pas pour décrire adéquatement une configuration spatiale, cas il existe en fait des nuances d'applicabilité. Pour en tenir compte, nous n'attribuons pas aux relations spatiales de simples valeurs de verités (cf. [André et al. 89]), mais plutôt un degré d'applicabilité. Ainsi peut on formuler des descriptions spatiales plus précises avec l'aide des différents degrés d'applicabilité définis par des modificateurs (cf. [Lakoff 73]) comme, par exemple, "directement", "immediatement", "juste", etc. En outre, le degré d'applicabilité est un critère de sélection supplémentaire au cas où plusieurs relations sont succeptibles de décrire une configuration spatiale.

On distingue trois grandes classes de relations spatiales³:

- 1. Relations topologiques ("dans", "chez", etc.) basées sur des concepts topologiques, comme par exemple la proximité ou l'intérieur.
- 2. Relations d'orientation ("à gauche", "à droite", etc.), qui dépendent de l'orientation de l'espace.
- 3. Relations décrivant des chemins et trajectoires d'objets ("le long", "à travers", etc.). Elles comprennent par là un aspect spatio-temporel.

Le calcul de l'applicabilité de relations spatiales d'orientation présuppose d'abord le choix de l'orientation de l'espace. La construction du système de référence dépend du mode d'utilisation de la relation (cf. [Hathout 89]).

On parle ainsi de l'utilisation intrinsèque d'une relation spatiale, quand l'orientation est fixée par l'organisation inhérente de l'entité correspondante. Les facteurs fixant les aspects inhérents d'une entité sont par exemple la direction de son mouvement (par ex. les voitures), la position des organes des sens (chez l'homme ou le robot) et son rôle fonctionnel (par ex. un bâtiment, un fauteuil).

Si au contraire on tient compte de l'orientation contextuelle induite par les caractéristiques internes d'un objet, il s'agira dans ce cas d'une utilisation extrinsèque: "Vue de l'immeuble, la banque se situe juste derrière l'église". Un cas particulier de l'utilisation extrinsèque est lorsque l'orientation est fixée par la position de l'observateur ("vue de ma/votre position"). On parle alors d'utilisation déictique d'une relation spatiale.

[Aurnague 91] et [Vieu 91] donnent un bon aperçu des approches existantes sur le traitement algorithmique de la sémantique des prépositions spatiales. Partant de nos propres travaux ([André et al. 87], [Wazinski 92]), nous avons élargi le modèle développé pour un espace bi-dimensionnel à une représentation graphique tri-dimensionnelle (cf. [Wazinski & Herzog 92]). Ainsi est rendu possible dans VITRA GUIDE l'utilisation des prépositions comme "sur", "sous", etc. De la même manière sera generalisée l'approche traitant les prepositions décrivant des chemins utilisée pour la première fois dans VITRA.

³De nombreuses méthodes de classification ont été proposées dans la littérature (cf. par ex. [Vandeloise 86], [Wunderlich & Kaufmann 90]). [Retz-Schmidt 88] en donne un bon résumé.

3 Présentation graphique d'informations spatiales

Dans la description habituelle d'itinéraires, on utilise plusieurs moyens, dont notamment les cartes géographiques, le plan d'une ville ou un atlas. Pour suivre un itinéraire par rapport à un systeme de référence défini par une représentation topologique d'un fragment du monde, l'interlocuteur (l'utilisateur du système) doit tout d'abord s'orienter. La perception visuelle joue un rôle déterminant pendant les phases d'orientation et de poursuite du parcours. Ces deux phases sont guidées par les caractéristiques du monde reel comme par exemple les marqueurs spatiaux, les points cardinaux, ou l'orientation naturelle de certaines plantes. L'homme peut ainsi utiliser son environnement réel ou une représentation topologique abstraite d'un fragment du monde. La représentation graphique reste cependant une manière naturelle pour l'homme lorsqu'il veut décrire des itinéraires.

Une représentation bi-dimensionnelle d'une carte permet la description d'un itinéraire à partir d'une perspective aérienne. Différents modes de présentation graphique peuvent être utilisés. On peut les classifier comme suit :

- 1. Cartes en 2D réalisées à partir d'une perspective aérienne.
 - (a) Mise en évidence graphique de l'itinéraire.
 - (b) Utilisation d'annotations sur certains marqueurs spatiaux.
 - (c) Aide à la recherche de chemins par des indications gestuelles.
- 2. Description à partir du point d'observation de l'interlocuteur
 - (a) Aide par des indications gestuelles.
 - (b) Animation graphique de ce que voit l'utilisateur au fil du chemin parcouru.

Un marquage complet de l'itinéraire par une mise en évidence graphique – comme par exemple dans le métro – permet d'obtenir une description minimale. En outre, un tel procédé peut être amélioré par la description en LN de certains marqueurs spatiaux. Des méthodes semblables sont appliquées dans les systèmes de navigation et d'information routière (cf. par ex. [Neukirchner 91]).

Ce procédé est approprié quand, après qu'on lui ait décrit l'itinéraire, la personne doit trouver toute seule son chemin. Certains aspects comme la facilité de reconnaissance et de vérification de la description par rapport au monde réel jouent un grand rôle dans ce cas (cf. [Hoeppner et al. 90]).

Une autre approche consiste à fournir à l'utilisateur du système une description incrementale. L'incrementalité dans ce contexte signifie que l'on donnera à l'utilisateur graduellement au moment opportun les informations nécessaires, comme le fait par exemple le passager dans une voiture.

En partant d'une représentation graphique bi-dimensionnelle, les indications gestuelles permettent une forme de description animée des itinéraires. Nous projetons d'intégrer dans VITRA le générateur de gestes ZORA ([Jung et al. 89]), qui utilise une main ou un doigt (voir figure 1).

L'utilisation habituelle de cartes 2D n'est cependant pas le moyen cognitif le plus adéquat pour présenter graphiquement un itinéraire. On utilise plus généralement la méthode déictique dans un espace tri-dimensionnel permettant une meilleure orientation. Notre première approche consiste en la présentation à partir d'un modèle tri-dimensionnel d'un point précis et de son environnement à un moment déterminé du parcours (voir figure 2). La combinaison de cette méthode avec la perspective aérienne en 2D permet à l'utilisateur de mieux reconnaître les marqueurs spatiaux sur l'itinéraire. Les indications gestuelles permettent de mieux montrer les changements de direction sur le parcours. Contrairement aux cartes 2D, il n'est pas exigé de l'utilisateur d'effectuer une transformation mentale de ces gestes par rapport à son système de référence.

L'utilisation de séquences d'images permet en plus de donner à l'interlocuteur une description plus détaillée de l'itinéraire à suivre. L' "animation graphique", qui était obtenue en 2D uniquement grâce aux gestes, prend içi une dimension beaucoup plus naturelle de par l'animation réelle des images. [Herzog 92] donne un aperçu des techniques de visualisation déjà réalisées dans VITRA.

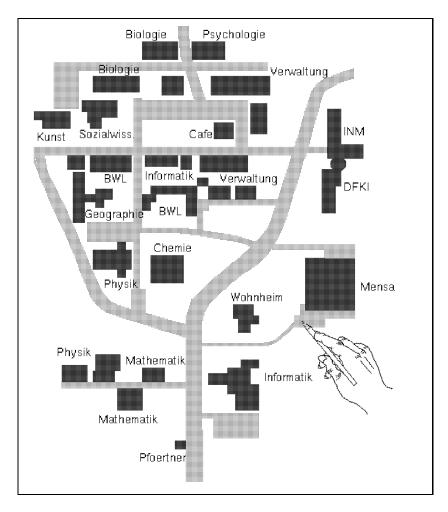


Figure 1: Indication gestuelle sur un plan

4 Intégration de la présentation linguistique et graphique

Un système d'informations multimodal comme VITRA GUIDE nécessite un langage de représentation des connaissances efficient et flexible. Conformément à la maxime de Grice sur la quantité (cf. [Grice 75]) — présenter autant d'informations que nécessaire, aussi peu que possible —, il est nécessaire de coordonner les modes de présentation disponibles.

Les récents systèmes d'informations⁴ utilisent différents modes de présentation comme le langage naturel, les méthodes graphiques, l'animation et les mondes virtuels. Le maintien de la cohérence et de la consistence des présentations multi-modales (cf. [Bandyopadhyay 91]) exige une utilisation coordonnée et concordante des différents modes. Une solution souvent utilisée est de concevoir la génération du texte de présentation comme un procédé de planification à but communicatif (cf. [André et al. 93], [Hovy 88]).

VITRA GUIDE utilise aussi cette approche de planification. Contrairement aux méthodes que nous

⁴Voir [Maybury 93] et [Sullivan & Tyler 91] pour un aperçu.

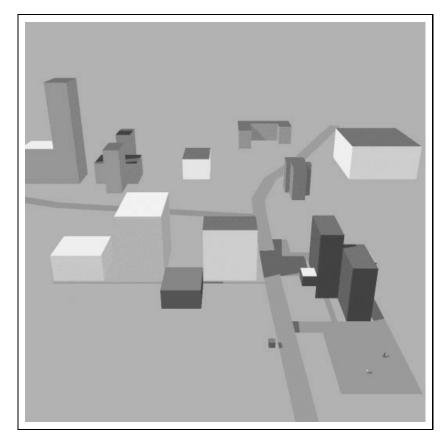


Figure 2: Modèle tri-dimensionnel du campus de l'université de Sarrebruck

venons de présenter, nous tenons compte des aspects spatio-temporels à cause de la description incrémentale des chemins. Par exemple le procédé de planification doit être en mesure de produire des descriptions du genre

Bitte folgenden Sie der Straße für drei Minuten. "Suivez cette rue pendant trois minutes."

En plus de ce type d'informations temporelles quantitatives, nous utilisons des relations temporelles qualitatives comme tout de suite ou après.

Bitte biegen Sie gleich rechts ab. "Tournez tout de suite à droite."

Des plans élémentaires seront produits par la division recursive des buts de planification en utilisant les morceaux de chemins définis par l'algorithme de recherche d'itinéraires. Ces plans vont être traités, suivant le mode de présentation, par la composante graphique ou linguistique.

Il faut tenir compte en particulier de la référentialité des objets dans les différents modes de présentation. Par exemple, si un marqueur spatial est mis en relief sur une carte, il faudra faire référence à cet attribut lors de la génération du texte en langage naturel. Par conséquent, le procédé de planification produit des références croisées entre les différents modes de présentation (cf. [Wazinski 92]).

La facilité de combinaison de ces différents modes dépend de l'application concrète. Par exemple, on n'utilisera la méthode graphique dans le domaine de la navigation automobile que partiellement, alors qu'un système d'informations urbain combinera la présentation graphique et textuelle.

5 Conclusion

Dans un système de génération automatique de descriptions d'itinéraires, la transmission d'informations spatiales indispensables à la description n'est pas restreinte à une description en langage naturel ou à une présentation graphique. On obtient un meilleur resultat pour l'utilisateur du système en combinant les deux méthodes.

La réalisation d'un système expérimental pour la génération de descriptions multi-modales d'itinéraires est le but de notre travail. Nous avons exposé les deux aspects de la présentation d'informations spatiales, à savoir d'une part la production d'expressions référentielles spatiales, et d'autre part, la présentation graphique. Des idées pour une intégration de ces différents modes de présentation ont aussi été présentées.

Les résultats déjà acquis sur la sémantique des prépositions spatiales, sur les techniques de visualisation, sur la génération d'indications gestuelles et sur la synthèse de documents multi-modaux constituent une base solide pour le développement futur du système VITRA GUIDE.

Remerciements

Nous tenons à remercier Alassane Ndiaye et Pierre Sablayrolles dont les suggestions et corrections ont permis de donner à ce papier une clarté et une lisibilité accrues.

Références

- [André et al. 87] E. André, G. Bosch, G. Herzog, and T. Rist. Coping with the Intrinsic and the Deictic Uses of Spatial Prepositions. In: K. Jorrand and L. Sgurev (eds.), Artificial Intelligence II: Methodology, Systems, Applications, pp. 375-382. Amsterdam: North-Holland, 1987.
- [André et al. 89] E. André, G. Herzog, and T. Rist. Natural Language Access to Visual Data: Dealing with Space and Movement. Report 63, Universit des Saarlandes, SFB 314 (VITRA), Saarbrücken, 1989. Presented at the 1st Workshop on Logical Semantics of Time, Space and Movement in Natural Language, Toulouse, France.
- [André et al. 93] E. **André**, W. **Finkler**, W. **Graf**, T. **Rist**, A. **Schauder**, and W. **Wahlster**. WIP: The Automatic Synthesis of Multimodal Presentations. In: M. T. Maybury (ed.), Intelligent Multimedia Interfaces, pp. 75–93. Menlo Park, CA: AAAI Press, 1993.
- [Aurnague 91] M. Aurnague. Contribution à l'Étude de la Sémantique Formelle de l'Espace et du Raisonnement Spatial: La Localisation interne en Français, Sémantique et Structures Inférentielles. Thèse de doctorat, Université Paul Sabatier, Toulouse, 1991.
- [Bandyopadhyay 91] S. **Bandyopadhyay**. Towards an Understanding of Coherence in Multimodal Discourse. Technical Memo TM-90-01, DFKI GmbH, Saarbrücken, 1991.
- [Davis 89] J. R. **Davis**. Back Seat Driver: Voice Assisted Automobile Navigation. PhD thesis, Media Arts and Science Section, MIT, Cambridge, MA, 1989.

- [Grice 75] H. P. Grice. Logic and Conversation. In: P. Cole and J. L. Morgan (eds.), Speech Acts, pp. 41-58. London: Academic Press, 1975.
- [Habel 87] C. Habel. Prozedurale Aspekte der Wegplanung und Wegbeschreibung. LILOG-Report 17, IBM, Stuttgart, 1987.
- [Hathout 89] N. **Hathout**. Un Modèle Logique pour le Raisonnement Spatial. Mémoire de DEA, Univ. Paul Sabatier, Toulouse, 1989.
- [Herzog et al. 89] G. Herzog, C.-K. Sung, E. André, W. Enkelmann, H.-H. Nagel, T. Rist, W. Wahlster, and G. Zimmermann. Incremental Natural Language Description of Dynamic Imagery. In: C. Freksa and W. Brauer (eds.), Wissensbasierte Systeme. 3. Int. GI-Kongreß, pp. 153-162. Berlin, Heidelberg: Springer, 1989.
- [Herzog 92] G. **Herzog**. Visualization Methods for the VITRA Workbench. Memo 53, Universität des Saarlandes, SFB 314 (VITRA), 1992.
- [Hoeppner et al. 90] W. Hoeppner, M. Carstensen, und U. Rhein. Wegauskünfte: Die Interdependenz von Such- und Beschreibungsprozessen. In: C. Freksa und C. Habel (Hrsg.), Repräsentation und Verarbeitung räumlichen Wissens, pp. 221-234. Berlin, Heidelberg: Springer, 1990.
- [Hovy 88] E. H. Hovy. Generating Natural Language under Pragmatic Constraints. Hillsdale, NJ: Lawrence Erlbaum, 1988.
- [Jung et al. 89] J. Jung, A. Kresse, N. Reithinger, und R. Schäfer. Das System ZORA: Wissensbasierte Generierung von Zeigegesten. In: D. Metzing (Hrsg.), GWAI-89, pp. 190-194. Berlin, Heidelberg: Springer, 1989.
- [Klein 83] W. Klein. Deixis and Spatial Orientation in Route Directions. In: H. L. Pick and L. P. Acredolo (eds.), Spatial Orientation: Theory, Research, and Application, pp. 283-311. New York, London: Plenum, 1983.
- [Lakoff 73] G. Lakoff. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. Journal of Philosophical Logic, 2:458-508, 1973.
- [Ligozat 92] G. **Ligozat**. Strategies for Route Description: An Interdisciplinary Approach. In: Spatial Concepts: Connecting Cognitive Theories with Formal Representations, Workshop Notes, 10th European Conf. on Artificial Intelligence, Vienna, Austria, 1992.
- [Maybury 93] M. T. Maybury (ed.). Intelligent Multimedia Interfaces. Menlo Park, CA: AAAI Press, 1993.
- [Müller 89] S. **Müller**. CITYGUIDE: Wegauskünfte vom Computer. Memo 37, Universität des Saarlandes, SFB 314 (VITRA), Saarbrücken, 1989.
- [Neukirchner 91] E. P. **Neukirchner**. Fahrerinformations- und Navigationssystem. Informatik Spektrum, 14(2):65-68, 1991.
- [Rauh & Schweitzer 87] A. Rauh und M. Schweitzer. Untersuchungen zur Verwendung scriptartiger Wissensrepraßentationen und Story Grammars bei der Generierung von Wegbeschreibungen. Paper B 158, L.A.U.D., Duisburg, 1987.
- [Retz-Schmidt 88] G. Retz-Schmidt. Various Views on Spatial Prepositions. AI Magazine, 9(2):95– 105, 1988.

- [Sullivan & Tyler 91] J. W. Sullivan and S. W. Tyler (eds.). Intelligent User Interfaces. New York: ACM press, 1991.
- [Vandeloise 86] C. Vandeloise. L'Espace en Français. Paris: Seuil, 1986.
- [Vieu 91] L. Vieu. Sémantique des Relations Spatiales et Inférences Spatio-temporelles: Une Contribution à l'Étude des Structures Formelles de l'Espace en Langage Naturel. Thèse de doctorat, Université Paul Sabatier, Toulouse, 1991.
- [Wahlster 87] W. Wahlster. One Word Says More Than a Thousand Pictures. On the Automatic Verbalization of the Results of Image Sequence Analysis Systems. Report 25, Universität des Saarlandes, SFB 314 (VITRA), 1987.
- [Wazinski & Herzog 92] P. Wazinski and G. Herzog. Coping with Topological and Directional Relations Based on 3-Dimensional Geometric Representations. In: Spatial Concepts: Connecting Cognitive Theories with Formal Representations, Workshop Notes, 10th European Conf. on Artificial Intelligence, Vienna, Austria, 1992.
- [Wazinski 92] P. Wazinski. Generating Spatial Descriptions for Cross-modal References. In: 3rd Conf. on Applied Natural Language Processing, Trento, Italy, pp. 56-63. Association for Computer Linguistics, 1992.
- [Wunderlich & Kaufmann 90] D. **Wunderlich** und I. **Kaufmann**. Lokale Verben und Präpositionen semantische und konzeptuelle Aspekte. In: S. Felix, S. Kanngießer, und G. Rickheit (Hrsg.), Sprache und Wissen, pp. 223-252. Opladen: Westdeutscher Verlag, 1990.
- [Wunderlich & Reinelt 82] D. Wunderlich and R. Reinelt. How to Get There From Here. In: R. J. Jarvella and W. Klein (eds.), Speech, Place, and Action, pp. 183-201. Chichester: Wiley, 1982.