

From Context to Sentence Form

Sabine Geldof

Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels
sabine@arti.vub.ac.be

Abstract

When generating utterances, humans may choose among a number of alternative sentence forms expressing the same propositional content. The context determines these decisions to a large extent. This paper presents a strategy to allow for such context-sensitive variation when generating text for a wearable, advice giving device. Several dimensions of context feed a model of the hearer's attention space, which, in terms of Information Structure Theory, determines the form of the sentence to be generated.

1 Introduction

When generating utterances, humans may choose among a number of alternative sentence forms expressing the same propositional content. Consider the following examples:

1. Amanda Huggenkiss proposes to meet you to talk about 'agents'.
2. Amanda Huggenkiss, she proposes to meet you to talk about 'agents'.
3. 'Agents' is proposed as a subject for a meeting, by Amanda Huggenkiss.

Discourse pragmatics research, mainly in view of NL understanding, reveals a loose coupling between a range of grammatical markers (morpho-syntax, word order and also prosody) and difficult to verify mental categories such as 'given' and 'new'. While unpredictability seems an inherent property of pragmatic phenomena (Verschuere, 1995) we want to experimentally introduce the observed variability in an NLG device to investigate its communicative effect. Our practical goal is to enhance the effectiveness of a wearable device that provides spoken advice to a user operating in a real-world physical environment. Given a particular pragmatic context, one or another formulation is more appropriate w.r.t. communicative success (Hovy, 1990). We focus on the hearer's context (as perceived by the speaker). Like Klabunde and Jansche (1998), we are interested in *linguistic form variations* related to information

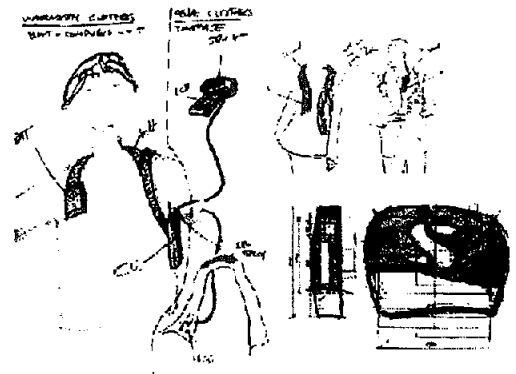


Figure 1: COMRIS Parrot design and prototype

packaging (Lambrecht, 1994) as an important aspect of addressee tuning. Taking into account *multiple context dimensions* acquired in real-time distinguishes our approach, also from other NLG research concerned with user adaptation, where only discourse and/or user profile are considered (e.g. (Krahmer and Theune, 1998; Paris, 1988)) or time and space from a linguistic-theoretical perspective (Maybury, 1991).

The work reported here is part of the COMRIS¹ project. In an information-rich situation (e.g. when visiting a conference), a user receives relevant advice (e.g. about interesting talks, interesting persons in the neighbourhood) from her 'parrot' (see figure 1)².

Related research issues in COMRIS are physical context sensing and information focusing through agents' competition for the attention of the user (Van de Velde et al., 1998). Context-sensitive text generation contributes to the latter and depends on the former. We earlier investigated how context determines the text generation process at the level of word choice (Geldof, 1999b). We proposed a multi-dimensional context model, encompassing discourse

¹COMRIS: Co-habited Mixed Reality Information Spaces (<http://arti.vub.ac.be/~comris/>)

²Reproduced with permission from Starlab nv (<http://www.starlab.org>)

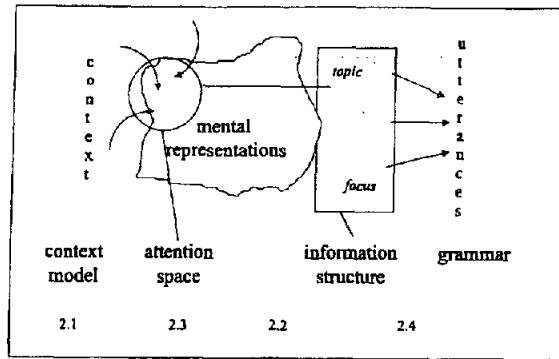


Figure 2: Suite of models corresponding to the context-sensitive NLG process

history, physical context and user profile (Geldof, 1999a). Real-time information about these different perspectives annotates the input structure of the template-based NLG component. We use TG/2, a rule-based engine that covers the continuum between templates and syntactic generation (Busemann, 1996). Making abstraction from planning and multi-sentential discourse phenomena allows us to focus on the subject of our research: context sensitivity and surface form. In this paper, we want to uncover how context affects the structure of utterances (vs lexical choice).

Section 2 presents the different steps of our approach: context modeling (2.1), information structure analysis (2.2), applied discourse pragmatics (2.3) and NLG strategy (2.4). Section 3 illustrates these ideas through scenarios and we conclude (section 4) with a discussion of our work.

2 Approach: from context to sentence form via attention focus

Our goal is to obtain a less intrusive device through context sensitivity of the spoken output. The presupposition is that utterances anchored to the hearer's multidimensional context will require less cognitive effort to attend to. Our strategy is based on the discourse pragmatic account for grammatical differences between utterances expressing the same propositional content. Figure 2 shows how we envision the connection between various disciplines relating context to utterance form.

Context is considered to activate particular mental representations in the hearer's mind (modeled as her attention space). In order to be communicative, the speaker hypothesises about this attention space and structures his utterance accordingly. Information Structure Theory accounts for this adaptation process. We use our earlier context model and developed a strategy for determining topic and focus based on the analysis of COMRIS' discourse prag-

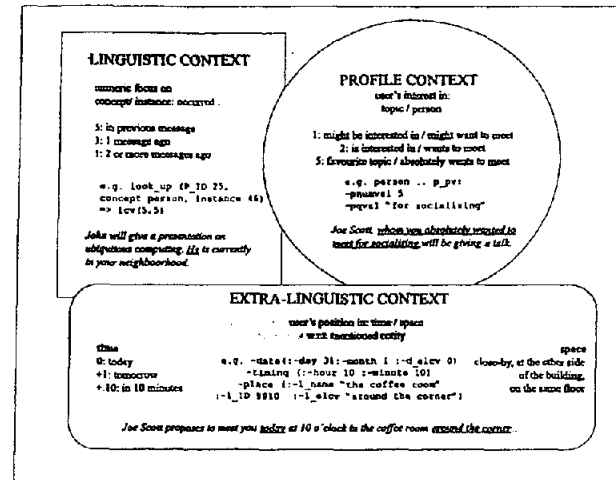


Figure 3: Overview of the different context perspectives in COMRIS' context model

matic situation.

2.1 Context modeling

Context perception and adaptation are important in research on wearable technologies. Nomadic Radio (Sawhney and Schmandt, 1999) adapts information delivery (news, email, phone-calls) to the user's physical environment through varying output sound technology. Remembrance agents act like memory extensions by pointing to related information in appropriate contexts (De Vault et al., 2000). Neither use linguistic form variation. Our I3 sister project, HIPS (Benelli et al., 1999) does and focuses on the interaction between physical and hypertext navigation for a wearable museum guide. Schmidt and colleagues provide a platform for inferring the relationship between low-level contextual data and application dependent concepts (Schmidt et al., 1999). When dealing with content delivery to human users, the use and interpretation of symbolic data in combination with quantitative data remains an important issue on the research agenda. Our context model is a first proposal in that direction.

When focusing on lexical phenomena like deictic expressions (this afternoon, here) and anaphora (she, the same topic) or the inclusion of appositives related to the hearer's profile (one of your favourite topics), we proposed a three-dimensional context model (see figure 3) in order to generate truly context sensitive expressions. Objects mentioned to the user are recorded in a discourse model, her location in space and time is monitored via beacons. The Information Layer provides user profile information (in terms of persons and topics of interest). Entities in the NLG input structure are annotated with contextual information of these different perspectives.

We will use the same multi-dimensional context model for building an attention space model of the hearer. Only for the physical context, we need additional reasoning on the time and location indexes in terms of the activity of the user (cfr. 2.3). Indeed, knowing which kind of activity the user is involved in at each moment (i.e. the ontology instances involved in that activity) we hypothesise on which person and keyword the user's attention is focused on.

2.2 Attention focus and Information Structure Theory

Other researchers have investigated attention focus in larger spans of discourse (McCoy and Cheng, 1991; Grosz and Sidner, 1986) and in dialogue (Jokinen et al., 1998). Corpus analysis (Rats, 1996) confirms the existence of a mechanism called topic, through which interlocutors strive at discourse coherence to reduce the cognitive effort of the hearer. The terminology used in the different frameworks is confusing, even contradictory (Bosch and van der Sandt, 1999). Information Structure Theory (Lambrecht, 1994) accounts for exactly those phenomena we are interested in: grammatical differences between allo-sentences (expressing the same semantic content). Lambrecht considers information structure as an integral part of the grammar of natural languages. After determining what to say, a speaker structures this information in terms of his own presupposition of the hearer's attention state. Identifiability (whether a shared representation exists in both interlocutors' minds) and activation status (how much a known representation is at the forefront of the hearer's mind (Chafe, 1987)) determine pragmatic role assignments. Topic, the role of aboutness is attributed to a discourse referent that is identifiable and more or less active. Focus is the unpredictable part of the utterance. Whereas all utterances have a focus (in order to be communicative), some may be topic-less. Lambrecht distinguishes 3 types of sentence constructions (according to whether the predicate, the argument or the whole sentence is in focus³) and demonstrates through grammatical analysis, that the first construction is the most natural one. Languages use different grammatical markers to realise information structure and there is no one-to-one correspondence between grammatical markers (e.g. definiteness, pronominalization, accentuation) and topic or

³Examples taken from (Lambrecht, 1994): ((SMALL CAPS indicate prosodic accent)

- (a) predicate focus: what did the children do? The children went to SCHOOL.
- (b) argument focus: who went to school? The CHILDREN went to school.
- (c) sentence focus (topic-less): what happened? The CHILDREN went to SCHOOL.

focus. In English, topic is preferably realised as an unaccented pronoun, while focus elements usually carry prosodic accent⁴.

2.3 COMRIS discourse pragmatics

There is no content-based discourse planning in COMRIS. The propositional content of parrot messages is provided by agents that represent particular user interests in the virtual world. A mechanism of competition for attention determines whether a message will actually be pushed to the user. As a consequence, the sentences to be generated are topic-less: each message conveys only new information, as if answering the hypothetical question: 'what is appropriate for me to do now?'. Thus they bare the danger of coming 'out of the blue', as in the following sequence:

- "There will be an interesting presentation by Amanda Huggenkiss about 'knowledge systems and AI'." (propagandist message)
- "Enric Plaza proposes to meet you to discuss about 'machine learning'." (appointment proposal)
- "Josep Arcos, who shares your interest in 'agents', is currently in your neighbourhood." (proximity alert)
- "Please note you have to give a presentation on 'NLG and context' within 5 minutes." (commitment reminder).

The intuition that such a sequence is not ideal from the communicative point of view, confirms our interpretation of information structure theory in view of communicative effect. Whereas topic expression creates discourse continuity (i.e. links the message to the context in a broad sense: an active mental representation), topic-less sentences can be assumed to require a higher cognitive effort from the hearer. Therefore our communicative strategy for COMRIS will be to look for a topic candidate within a given propositional content. To be communicatively more effective, we try to somehow link a message to the attention space of the user-hearer.

Obviously, the hearer's mind is a black box and all we can do is hypothesise about the activation of mental representations by contextual factors. In line with our previous work, we argue that the 3 dimensions of the user's context (linguistic, physical and profile) should be taken into account. Given the COMRIS ontology, the attention state model can be represented as a simple tree structure (see examples in section 3): each utterance conveys information

⁴This is a simplification of Lambrecht's analysis. Our point is that less prosodic accents reduce the cognitive effort of the hearer, which is our goal. Combined with the choice of sentence structure, it constitutes our strategy for reduced obtrusiveness.

about an event characterised by a key-word (-list), involving a person, and possibly a time/location specification. Thus we will search in the hearer's discourse and physical context which are the activated instances of the concepts event, person, key-word and time/location. To find out which instances are contributed by the physical context, we hypothesise about the user's current activity by comparing her physical position with the conference programme or her agenda. For instance, if we know that the user is attending a particular presentation, we can query the conference program for the speaker and the keywords of that presentation. Alternatively, if the user's physical location confirms that she attends an appointment, her agenda will reveal the name of the person she's meeting and maybe some topics of discussion. Any of these instances may also carry context annotation w.r.t. the user's interest profile. Section 3 explains this further through scenarios.

2.4 NLG strategy

Assignment of topic and focus follows from our application of Information Structure Theory to the discourse pragmatic situation in COMRIS. Our search for a topic candidate in the NLG input structure considers time pressure first, then the activation of entities via discourse or activity and finally the hearer's interest profile, as detailed in the following rules:

1. (*physical context first*) If the NLG input structure contains a time expression that is annotated as being very close to the current point in time (physical context value), then let the time expression be the topic, opening the sentence and carrying a prosodic accent. The sentence structure is predicate focus.

e.g. Please note that, within FIVE MINUTES, you have to give a presentation on 'NATURAL LANGUAGE GENERATION AND CONTEXT'.

2. (*topic candidate in attention space*) If one of the entities of the input structure is also present in the (hearer's) attention space map, let it be the topic, realised as an unaccented pronoun (preferred topic marking) in case it occurred in the immediate linguistic context⁵ or as a left-dislocated constituent in case it was present in the physical context.

e.g. She will give a presentation on KNOWLEDGE SYSTEMS and AI.

AMANDA HUGGENKISS, she will give a presentation on KNOWLEDGE SYSTEMS and AI.

3. (*profile context also matters*) If none of the above situations occur, verify whether any of the entities of the input structure has a high

⁵immediately preceding message, not more than X ago, where X is a time threshold

profile value (indicating the hearer's interest in that keyword or person). If the physical context also allows topic shift, use an argument focus structure (after introducing the new topic):

e.g. Someone interested in 'AGENTS' is in your neighbourhood. It's JOSEP ARCOS⁶.

4. Else (default) use a sentence focus structure.
e.g. PENN SILL proposes to MEET you to talk about 'NLG and CONTEXT'.

The scenarios below will further concretise the relationship between context, attention space and topic-focus assignment, but the above examples already illustrate our main point. The first 3 rules are aimed at linking an element of the propositional content to the user's attention focus, in virtue of the preceding discourse, the physical context or her interest profile. Topic expression often leads to de-accentuation. In other words, rule 4 applies when there is no way to anchor the utterance to the user's context and requires to accent every information entity. Empirical experiments will have to verify the hypothesis that the non-default sentence constructions are perceived as less intrusive.

3 Scenarios

This section illustrates (with examples from simulated conference data) how the attention space is derived from the context and how rules for topic/focus assignment are applied. In each scenario the previous utterance of the parrot to the user -if recent enough-, constitutes the linguistic context (lcv), the user's current activity activates entities via the physical context (elcv). The tree diagram shows the corresponding attention space map. The propositional content (input to the NLG process) consists of the message type, an instance of event, person, keyword(s) and possibly time expression. Finally we compare the context sensitive NLG output with the default output.

3.1 Scenario 1: topic-focus structure

At the moment of utterance, the hearer's context can be characterised as follows:

linguistic context: "There will be an interesting presentation on 'knowledge systems and AI' by Amanda Huggenkiss, this afternoon."

physical context: user is attending a presentation on 'machine learning' by Penn Sill.

This situation may be analysed as activating discourse referents in the hearer's mind as represented in the attention space map of figure 4.

⁶Note that, in case the object marked for high interest is the person, a more abbreviated sentence construction is appropriate: 'Josep Arcos is in your neighbourhood'. Since the user indicated herself that she is interested in this person, no need to further characterise him.

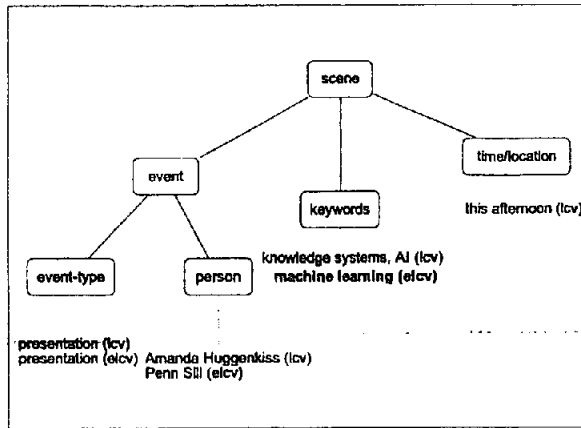


Figure 4: Attention Space map for scenario 1, built from linguistic (lcv) and extra-linguistic context (elcv)

propositional content: *appointment proposal: Enric Plaza, machine learning.* One of these entities also appears in the attention space: machine learning.

This situation leads to the application of rule 2: 'machine learning' will be assigned the role of topic, while other entities of the input structure ('appointment proposal' and 'Enric Plaza') will receive the role of focus. This yields the following output:
 "MACHINE LEARNING, it's also the subject of an APPOINTMENT PROPOSAL by ENRIC PLAZA."
 Compare with the default sentence construction:
 "ENRIC PLAZA proposes an APPOINTMENT to talk about MACHINE LEARNING."

3.2 Scenario 2: topic shift

linguistic context: "There will be an interesting presentation on 'knowledge systems' and 'AI' by Amanda Huggenkiss, this afternoon."

physical context: user is leaving a presentation, on her way to a meeting with Richard Benjamins on Machine Learning.

This situation leads to the attention space map of figure 5.

propositional content: *proximity alert, Josep Arcos, agents (p.pv, 5).* The profile value annotation indicates that this keyword is of high interest to the user (as indicated by herself, e.g. at conference registration).

The physical context is such that it allows for a shift of topic (user is not yet talking to Richard Benjamins), which makes rule 3 applicable: 'agents' will be introduced as a new topic, followed by an argument focus structure:

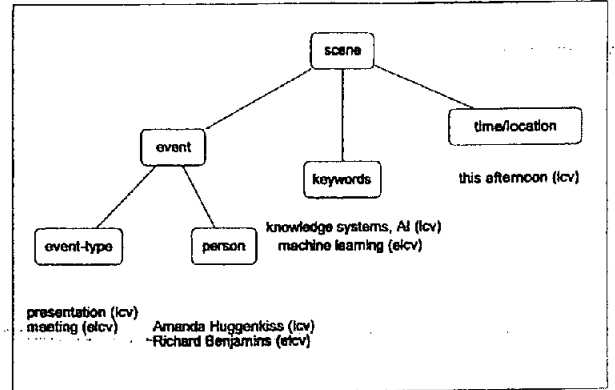


Figure 5: Attention Space map for scenario 2

"Someone interested in 'AGENTS' is close to you: it's JOSEP ARCOS."

Compare to the default expressions:

"JOSEP ARCOS, who's interested in 'AGENTS', is close to you."

4 Discussion

In this paper we proposed an NLG strategy that relates aspects of the hearer's multidimensional context to grammatical variations of sentences. The interpretation of the COMRIS pragmatic situation in terms of Information Structure Theory leads to decreasing the hearer's cognitive effort by linking the propositional content to her broad context. This is marked grammatically in the resulting utterance. Although the details may be application dependent, we believe the general idea holds for context-aware verbal interaction on wearable devices. Experiments in another application area would involve the elaboration of another ontology and might reveal other grammatical markers. We see some limitations and challenges for further research. The approach critically depends on progress in context capturing and especially its high-level interpretation (Schmidt et al., 1999). The use of more sophisticated AI techniques could account for the uncertainty involved in attention space modeling and the indeterminism in mapping pragmatic features to grammatical markers. As more hardware and software becomes available and integrated towards the end of the COMRIS project, we plan to perform real-world experiments. We can already evaluate our strategy by comparing results from generation with and without considering context, the former producing more varying and more natural output. Our major contribution consists in linking work on focus of attention to real-time monitoring and modeling of different hearer context dimensions and in providing a framework for experimentation and elaboration of NLG techniques for the interaction devices of the future: wearables.

Acknowledgments We appreciated the support of Stephan Busemann w.r.t. the use of TG/2. Walter Van de Velde, Kristiina Jokinen and Jacques Terken provided interesting feedback on the ideas developed here. Many thanks also to the partners of the COMRIS project who are concerned with integrating the different modules, especially Stefan Haustein and Ronald Schroten. This work is funded by the EU LTR research project COMRIS (LTR-25500) within the framework of I3 (Intelligent Interaction Interfaces).

References

- G. Benelli, A. Bianchi, P. Marti, E. Not, and D. Sen-nati. 1999. Hips: Hyper-interaction within physical space. In *Proc. of IEEE'99*, Florence.
- P. Bosch and R. van der Sandt, editors, 1999. *Focus: Linguistic, Cognitive & Computational Perspectives*, chapter Preface. Studies in natural language processing. Cambridge University Press, Cambridge.
- S. Busemann. 1996. Best-first surface realization. In D. Scott, editor, *Proc. of the 8th Intl. workshop on Natural Language Generation*, pages 101-110, Herstmonceux Castle. University of Brighthon, UK.
- W. Chafe. 1987. Cognitive constraints on information flow. In R.S. Tomlin, editor, *Coherence and grounding in discourse*, pages 21-50. John Benjamins B.V.
- R. De Vaul, B. Clarkson, and A.Sandy Pentland. 2000. The memory glasses: towards a wearable, context aware, situation-appropriate reminder system. In *Preprints of the CHI-2000 workshop on Situated interaction in ubiquitous computing*, number 2000-7 in Interner Bericht, Computer Science Department, Karlsruhe, Germany.
- S. Geldof. 1999a. Parrot-talk requires multiple context dimensions. In *Proc. of the 2nd Intl. and Interdisciplinary Conference on CONTEXT*, number 1688 in LNAI, pages 467-470. Heidelberg. Springer Verlag.
- S. Geldof. 1999b. Templates for wearables in context. In T. Becker and S. Busemann, editors. "May I speak freely?" *Between templates and free choice in natural language generation. Workshop at the 23rd German Annual Conference for AI (KI99)*. DFKI-D-99-01, Saarbrücken. DFKI.
- B.J. Grosz and C.L. Sidner. 1986. Attention, intentions and the structure of discourse. *Computational Linguistics*, 12(3):175-204.
- E.H. Hovy. 1990. Pragmatics and natural language generation. *Artificial Intelligence*, 43(2):153-198.
1998. *Proc. International Workshop on Natural Language Generation*. Niagara-on-the-lake, Canada.
- K. Jokinen, H. Tanaka, and A. Yokoo. 1998. Planning dialogue contributions with new information. In INLG98 (INL, 1998), pages 158-167.
- R. Klabunde and M. Jansche. 1998. Abductive reasoning for syntactic realisations. In INLG98 (INL, 1998), pages 108-117.
- E. Krahmer and M. Theune. 1998. Context-sensitive generation of descriptions. In *Proc. of the 5th Intl. Conf. on Spoken language processing (ICSLP'98)*, pages 1151-1154.
- K. Lambrecht. 1994. *Information structure and Sentence Form*. Cambridge University Press, New York (NY).
- M. Maybury. 1991. Topical, temporal and spatial constraints on linguistic realization. *Computational Intelligence*, 7:266-275.
- K.F. McCoy and J. Cheng. 1991. Focus of attention: constraining what can be said next. In C.L. Paris, W.R. Swartout, and W.C. Mann, editors, *Natural language generation in Artificial Intelligence and Computational Linguistics*, pages 103-124. Kluwer Academic Publishers, Norwell, MA.
- C.L. Paris. 1988. Tailoring object descriptions to a user's level of expertise. *Computational Linguistics, special issue on User Modeling*, 14(3):64-78, September.
- M. Rats. 1996. *Topic management in information dialogues*. PhD dissertation, Katholieke Universiteit Brabant, Tilburg, NL.
- N. Sawhney and C. Schmandt. 1999. Nomadic radio: Scaleable and contextual notification for wearable audio messaging. In *Proc. of CHI'99, the ACM SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, Pennsylvania.
- A. Schmidt, K.A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, and W. Van de Velde. 1999. Advanced interaction in context. In H. Gellersen, editor, *Proc. of Intl. workshop on Handheld and Ubiquitous Computing (HUC99)*, number 1707 in LNCS, Heidelberg, Germany. Springer Verlag.
- W. Van de Velde, R. Schrooten, and S. Geldof. 1998. Competition for attention. In M.P. Singh, A.S. Rao, and M.J. Wooldridge, editors. *Proc. of ATAL: 4th Int. Workshop on Agent Theories, Architectures and Languages*, volume 1365 of LNAI, pages 282-296. Springer Verlag, Heidelberg.
- J. Verschuere. 1995. The pragmatics perspective. In J. Verschuere, J.O. Oestman, and J. Blommaert, editors, *Handbook of Pragmatics*, pages 1-19. John Benjamins Publishing Co., Amsterdam.