This is the accepted manuscript (after review) of an article published by
Taylor & Francis in December 2010 in Applied Artificial Intelligence 24(10):914–952.
The published version of this paper is available online at
http://doi.org/10.1080/08839514.2010.529258

MARQUIS: Generation of User-Tailored Multilingual Air Quality Bulletins

Leo Wanner[1,2], Bernd Bohnet[3], Nadjet Bouayad-Agha[2],
François Lareau[4], Daniel Nicklaß[5]

[1]Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain

[2]Department of Communication and Information Technologies,
Pompeu Fabra University, Barcelona, Spain

[3]Institute for Natural Language Processing (IMS), University of Stuttgart, Germany

[4]Department of Computing, Macquarie University, Sydney NSW 2109, Australia

[5]Autodesk Consulting Geospatial, Ludwigsburg, Germany

Abstract

Air pollution has a major influence on health. It is thus not surprising that air quality (AQ) increasingly becomes a central issue in the environmental information policy worldwide. The most common way to deliver AQ information is in terms of graphics, tables, pictograms or color scales that display either the concentrations of the pollutant substances or the corresponding AQ indices. However, all of these presentation modi lack the explanatory dimension; nor can they be easily tailored to the needs of the individual users. MARQUIS is an AQ information generation service which produces user-tailored multilingual bulletins on the major measured and forecasted air pollution substances and their relevance to the human health in five European regions. It incorporates modules for the assessment of pollutant time series episodes with respect to their relevance to a given addressee, for planning of the discourse structure of the bulletins and the selection of the adequate presentation mode, and for generation proper. The positive evaluation of the bulletins produced by MARQUIS by users shows that the use of automatic text generation techniques in such a complex and sensitive application is feasible.

Keywords: air quality, multilingual report generation, document planning, linguistic generation, numeric time series

1 Introduction

To be adequately informed about air quality (AQ) is essential. It is hazardous to do outdoor sports in times and areas of elevated ozone, carbon monoxide, or particulate matter concentrations. Individuals with cardiovascular and respiratory conditions suffer from a mere exposure to high ozone or particulate matter concentrations, and children are very sensitive to all kinds of air pollution. Therefore, AQ is increasingly a central issue of the environmental information policy worldwide. Station networks across countries monitor the concentrations of air pollutant substances such as ozone (O_3), particulate matter (PM_{10} and $PM_{2.5}$), nitrogen monoxide and dioxide (NO_x), sulphur dioxide (SO_2), and carbon monoxide (CO). However, the question of how the monitored concentrations are to be turned into information for citizens received so far much less attention.

Up to date, the measured concentrations or indices thereof have most often been presented to the public in terms of tables, distribution curves, pictograms, or color scales. However, tables and curves are not self-explanatory. Citizens with no background on air pollution are not able to interpret them and draw the proper conclusions with respect to their behavior. Pictograms and color scales reflecting air quality or indices of individual pollutants are intuitively clear, but they do not provide any explanatory information, nor context-related and addressee-tailored health warnings or advice. New generation intelligent AQ information services are needed! This need has already been voiced in several environmental forums; cf., e.g., (Peinel *et al.*, 2000; Johansen *et al.*, 2001; Bøhler *et al.*, 2002; Karatzas, 2007). The experts agree that such services must: (a) incorporate intelligent data interpretation needed to analyze the course of the measured concentrations and assess their relevance; (b) tailor their information to the needs of the users; (c) rely upon the textual mode as the central mode; and (d) offer the information via all modern communication channels.

A few prototypical AQ services that attempt to cover (c) and (d) have been developed so far; cf. (Busemann & Horacek, 1997; Bohnet *et al.*, 2001; Bøhler *et al.*, 2002). In this article, we describe the MARQUIS service, which attempts to equally cover (a) and (b) and push forward the state of the art in report generation with respect to (c) and (d). MARQUIS provides multilingual and multimodal air quality information for five European regions. Its most innovative features are:

(i) reference to a default user profile typology, with the option of a flexible individualization of each profile by the users;

(ii) coverage of the major modern communication channels (adapting the writing style to the channel in question): internet, email, mobile phone (SMS and WAP), TV and printed media;

(iii) coverage of the major air pollutant substances of each MARQUIS region;

(iv) advanced air quality forecasting models;

(v) user-tailored assessment of pollutant time series episodes with respect to their relevance to a given user and retrieval of complementary information from an external background knowledge base;

(vi) interpretation of measured pollutant concentrations, making reference to meteorological conditions that cause or influence observed air pollution and background knowledge;

(vii) advanced computational linguistics techniques for generation of multilingual material.

In the remainder of the article, we focus on the features that make MARQUIS an *intelligent* service suitable for daily operational use, concentrating on one mode (the text mode) and one communication channel (the web). Section 2 provides a short overview of the input from which the production of air quality bulletins in MARQUIS starts. Section 3 describes how user modelling and interaction of the service with the user has been addressed. In Section 4, the architecture of the MARQUIS service is briefly outlined. Sections 5 and 6 present the central parts of the service: the air quality interpretation module (Section 5) and the document planning and linguistic realization modules (Section 6). In Section 7, the results of the performance evaluation of the document planner and the linguistic generator are discussed, before, finally, in Section 8, some conclusions are drawn and some lines of future work are sketched.

2 Where Do We Start From?

The input for the production of AQ bulletins in MARQUIS comes from four different sources: (1) regional monitoring networks, which deliver measured (raw) time series for a number of pollutant substances, (2) AQ asessment models, which provide forecasts of pollutant concentrations and AQ index time series, (3) meteorological models, which supply measured and forecasted time series for meteorological conditions, and (4) a background knowledge base, which contains regional and user-specific AQ related information necessary for high quality user-tailored bulletins. That is, the MARQUIS application is a showcase application of report generation from numeric time series, complemented by background knowledge.

In the past, numeric time series have often been verbalized using Natural Language Generation (NLG). Consider, e.g., stock market evolution (Kukich, 1983), labour market statistics (Rösner, 1986; Iordanskaja *et al.*, 1992), weather reports (Goldberg *et al.*, 1994; Coch, 1998;

Sripada *et al.*, 2003), retail statistics (Iordanskaja *et al.*, 1992), and, more recently, gas turbine monitoring reports (Yu *et al.*, 2007) and medical intensive care unit monitoring bulletins (Portet *et al.*, 2009). As mentioned above, several generators also verbalize AQ data. The size of the AQ time series is comparable with the size of the meteorological time series. However, MARQUIS (unlike the previous AQ information generators) has also to cope with different types of time series, which require different interpretations. Some of these time series are derived from other series (as the AQI time series is derived from the individual pollutant time series), others correlate (as the time series of selected meteorological conditions correlates with the pollutant concentration time series), and certain values of a given time series are associated with background information.

2.1 Monitored Pollutant Concentration Time Series

As mentioned in the Introduction, the MARQUIS service covers five European regions: Baden-Württemberg (Germany), Catalonia (Spain), Finland, Portugal, and Upper Silesia (Poland). All of them have operational air pollution monitoring networks, which monitor several of the standard pollutant substances; cf. Table 1.[1]

PM_{10}, NO_2, O_3, SO_2 and CO are measured in nearly all regions—except in PT. The monitoring of $PM_{2.5}$ is still less common in Europe; from the five MARQUIS regions, only two (FIN and PT) monitor $PM_{2.5}$. Obviously, the relevance of the individual pollutants varies from region to region. Thus, in BW, CAT and PT, PM_{10} and O_3 are crucial, while, for instance (as already mentioned above), CO and SO_2 are of no relevance in PT. In contrast, in US, it is these two substances which are of special prominence.

The concentration of each pollutant substance is measured periodically at a pollutant-specific rate for each location at which a monitoring station is located. Table 2 shows an excerpt

[1]The network providers are: for BW the *Landesanstalt für Umwelt, Messungen und Naturschutz* (LUBW), for CAT the *Servei Meteorologic de Catalanya* (SMC), for FIN the *Finnish Meteorological Institute* (FMI), for PT the *Portuguese Environmental Institute*, and for US the *Institute for Ecology of Industrial Areas* (IETU).

Table 1: Spectrum of the monitored air pollutants

Region	Air pollutants					
	PM_{10}	$PM_{2.5}$	NO_2	O_3	SO_2	CO
Baden-Württemberg (BW)	×		×	×	×	×
Catalonia (CAT)	×		×	×	×	×
Finland (FIN)	×	×	×	×	×	×
Portugal (PT)	×	×	×	×		
Upper Silesia (US)	×		×	×	×	×

from the hourly pollutant concentration time series for the monitoring station Karlsruhe-Nordwest in BW on July 20th, 2006.

Table 2: Excerpt from the pollutant time series at Karlsruhe-Nordwest on July 20th, 2006

	Time									
Polutant	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00
O_3	13	61	57	47	26	44	78	84	129	164
PM_{10}	27	33	35	39	63	78	50	50	45	43
NO_2	78	49	47	55	78	64	43	48	27	12
SO_2	3,15	3,35	2,6	2,7	4,8	4,0	4,0	2,7	2,9	2,1

As the figures in the table indicate, the hourly concentrations of the individual pollutants may vary substantially, calling for interpretation and justification. A daily distribution curve of a pollutant measured at a specific location correlates with the context of measuring (the time, the day of the week, the type of surroundings, i.e., exposed to traffic or background, street canyon or flat landscape, etc.) and the meteorological conditions. Especially the latter may make the concentration rise to health threatening levels.

2.2 Quantitatively and Qualitatively Forecasted Air Pollution

Quantitative and qualitative forecast of air pollutant concentrations is already an essential part of the state-of-the-art AQ information services. It was thus crucial to integrate it into the MARQUIS service. Table 3 summarizes the available forecast spectrum in all five MARQUIS regions. It shows that in addition to the individual pollutants, for CAT and PT, the *Air Quality Index* (AQI) is being forecasted. AQI is a qualitative and more abstract reflection of the concentrations of all pollutant substances. The calculation of AQI is, as a rule, based on national or regional air quality guidelines; see Section 5.

Table 3: Spectrum of the forecasted pollutants

	Forecasted air pollutants					
Region	PM_{10}	NO_2	O_3	SO_2	CO	AQI
BW	×		×			
CAT			×			×
FIN	×	×	×		×	
PT	×	×	×			×
US	×	×	×	×	×	

Both the forecasted pollutant concentrations and the forecasted AQI are provided in terms of time series of the type presented in Table 2. Although a contrastive discussion of the models used for the derivation of these time series would be an interesting topic, it is far beyond the scope of this article—such that some cursory remarks must suffice.

For quantitative forecasting, MARQUIS uses local scale, i.e., regional, air quality forecasting models.[2] For PT, FIN, and CAT, models developed prior to MARQUIS (by the members

[2]At the first glance, the use of one of the available large scale forecasting models for all MARQUIS regions seems very attractive. Even more so, since large scale models are able to capture the long range transport of pollutant substances—which would have certainly been very beneficiary for the explanation of the measured

of the project consortium—as in the case of PT and FIN—or by external institutions—as in the case of CAT) are used. For FIN, the CAR-FMI model is used (Kukkonen *et al.*, 2001); for PT, we use a PREV'AIR type model for NO_2 and statistical models for PM_{10}, O_3 and the AQI (Ferreira *et al.*, 2000; Neto *et al.*, 2005). In CAT, a physico-chemical Lagrangian model is run for O_3 forecasting (Grell, 1993), and qualitative manual forecasting is done for AQI. In BW, O_3 forecasting in summer time is also manual; for the prediction of PM_{10}, two models are used: the PT model, which has been adapted for selected locations in BW, and a kNN machine learning-based model that has been developed explicitly for MARQUIS (Lohmeyer *et al.*, 2007). In US, the SINZAP model has been developed for forecasting the concentrations of all substances (Bronder *et al.*, 2007).

All models used in MARQUIS provide stable high quality forecasts necessary for a citizen-oriented AQ information service. For an exhaustive validation of the models, see the publications cited above.

2.3 Meteorological Conditions Time Series

Meteorological conditions are needed in MARQUIS, on the one hand, for the AQ forecasting models and, on the other hand, for user-tailored interpretation (and motivation) of given air pollutant concentrations. Among the monitored and forecasted meteorological conditions are, among others: precipitation, wind strength, wind direction, temperature, dewpoint, and humidity. The total number of measured conditions is similar to the number captured, e.g., in the SemTime-Mousam weather forecast generator (Sripada *et al.*, 2003).

The meteorological time series (both monitored and forecasted) are provided by the *Finnish Meteorological Institute* for BW, FIN, and US. For CAT, they are provided by the *Servei Meteorologic de Catalunya*, and for PT, by the *Portuguese Environmental Institute*. The

concentrations. However, the evaluation of the two most prominent European large scale models, EURAD (Jakobs *et al.*, 2005) and THOR (Van Loon *et al.*, 2004) revealed that they perform too poorly to be used in MARQUIS; for the evaluation, see (Lohmeyer *et al.*, 2007).

accuracy of the meteorological forecasting models is attested and will not be elaborated on in this article.

2.4 AQ Background Knowledge

In addition to the numeric monitored and forecasted air pollutant and meteorological time series, MARQUIS draws upon static background knowledge. This knowledge is mostly of regulatory and cultural nature. It is region-specific because the perception of AQ in general and of pollutant concentrations in particular varies from one European region to another. For instance, in Finland and Germany, citizens are much more sensitive to AQ issues than in Spain or Portugal. This is also reflected in the environmental regulations of each country. In the context of report generation, this concerns:

(i) the correspondence between AQI scales (e.g., 1 to 6) and qualitative ratings such as "good", "bad" and "satisfactory";

(ii) the correspondance between pollutants' concentrations and quantitative ratings such as "high", "low" and "moderate";

(iii) the correspondence between numeric concentrations and AQI gradients and ratings such as "unchanged", "slight", "strong", etc.;

(iv) the threshold concentrations for each pollutant substance and the AQI beyond which health warnings and calls for action must be issued;

(v) canned text messages concerning health risks and precautionary measures that are to be taken by the affected group of citizens when concentrations reach a given threshold.[3]

[3] Statements on the harmlessness of given pollutant concentrations, health risk warnings, calls for action (e.g., concerning traffic regulations) and the like are represented as canned text messages because they are legal statements spelled out in national and European laws, directives and regulations.

Our empirical studies also show that the subjective interpretation of a given AQI and a pollutant concentration with respect to their "communicative significance", i.e., whether they are high enough to be worth mentioning (although below any threshold), varies from region to region.

A further aspect of the static background knowledge concerns culture-specific lexicalization. Among others, for instance, the interpretation and naming of the time intervals of a day differs from region to region. Thus, the Spanish MAÑANA 'morning' extends more or less until 2 PM and the TARDE 'afternoon' until 8 or 9 PM. The German MORGEN 'morning' can go until 12:00, and the NACHMITTAG 'afternoon' until 5 PM at the latest. The diverging interpretation is occasionally reflected by the vocabulary; for instance, in German, a special term for 'time before noon'—VORMITTAG is available, while in other MARQUIS languages this time is still called MORNING.

The static background information is specified in tables for all regions and all languages covered by MARQUIS. The use of this information varies. Thus, the communicative significance table guides (among other parameters) content selection. The canned message tables are repositories of chunks of information, some of which are included into the document plan when certain constraints are fulfilled. The other tables are mapping tables.

3 Modelling the Information Needs of the Addressee

Most topics targeted by NLG call for a differentiation of the content, discourse structure and language style for different addressees. As a rule, these differentiations are directly or indirectly captured by a user model (Zukerman & Litman, 2001). A user model may consist of a list of conversation settings, as, e.g., in Hovy's PAULINE (Hovy, 1990), be predefined along the 'naïve user'–'expert' scale, as, in (Paris, 1993)—potentially with the option of incremental individualization, consist of the user's beliefs as in (Zukerman & McConachy, 1994) or be derived from a personal data record, as in (Cawsey *et al.*, 2000). Each parameter (or a combination thereof) in the user model usually serves as a criterion in the content selection task (in some implementations, also in the discourse structure and style determination tasks). Thus, Walker *et al.* (2004) base the selection of the content on restaurants to be offered to a user based on this user's rating of a set of preferences inquired before; Sripada *et al.* (2003) tailor weather forecasts to users depending on their location and tasks.

The evaluation of the personal data records of individuals interested in AQ information that have been gathered in a survey carried out within MARQUIS (Molina *et al.*, 2005) showed that the information needs of these individuals can often be grouped according to certain criteria. But, at the same time, the strong influence of air pollution on the health of the affected individuals may also require highly personalized information.[4] We solve this dilemma by a two-level user model. The first level is given by a rough typology of default *user profiles*; the second level consists in the personalization of the default profile by the user.

3.1 Default User Profile Typology in MARQUIS

For the definition of the default user profile typology in the context of AQ information, at least three dimensions come into play: (i) expertise with respect to air pollution, (ii) air pollution sensitivity of the target audience, and (iii) the preferred communication channel. The expertise dimension allows us to specify what kind of background information a user needs, to what extent the AQ information should be qualitative (and thus easier to understand) or quantitative (and thus require further interpretation), and in which mode the information is preferrably to be presented (text, table, or graphic)—provided the communication channel chosen by the user allows for variation of the mode. The air pollution sensitivity dimension lets us link the information delivery to specific AQ levels and concentrations of different pollutant substances, include or omit health warnings, mention or not the weather conditions, etc. The communication channel dimension determines further the mode of presentation, the conciseness

[4] In each of the five MARQUIS regions, about 100 individuals with different profiles have been interviewed.

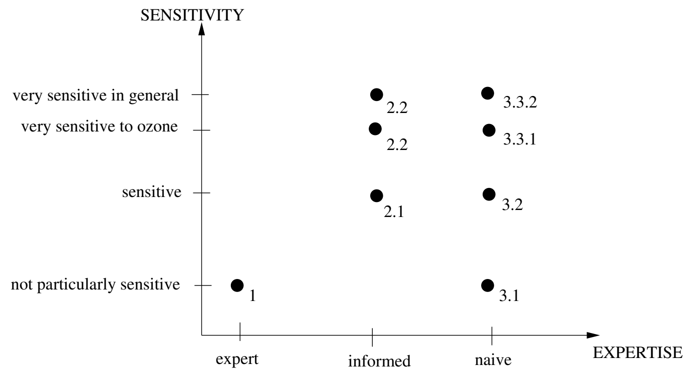


Figure 1: Coordinates for default user profiles

(i.e., the amount) of the information offered, etc. As mentioned in Section 1, in this paper, we focus on the web as the communication channel. A comprehensive discussion in the light of all other communication channels covered in MARQUIS (email, mobile phone (SMS and MMS), printed media, and TV) can be found in (Molina *et al.*, 2005).

The empirical study revealed that the expertise dimension should distinguish at least between 'expert', 'informed', and 'naive' users, and the sensitivity dimension between 'not particularly sensitive', 'sensitive', 'very sensitive with respect to ozone' and 'very sensitive in general'. Coordinate pairings in Figure 1 showed to be of significant relevance. Each pairing determines a default user profile (identified in Figure 1 by a number). Figure 2 displays the resulting user profile typology in which the numbers in Figure 1 are associated with telling names.

Thus, AQ experts predominantly desired to receive information at the level of detail corresponding to 'not particularly sensitive' addressees. Medical professionals requested to be 'informed' with respect to AQ and its influence on health; the coordinate on the sensitivity axis assigned to a medical professional depends on their specialization, i.e., on the profile of their patients: general medicine specialists are assigned the feature 'sensitive' (in a representative group of patients, there are always patients sensitive to bad AQ); heart specialists are assigned the feature 'very sensitive to ozone' (heart patients are particularly sensitive to this

1. domain professional
2. medical professional
   2.1 respiratory disease specialist
   2.2 heart specialist
   2.3 general medical professional
3. public
   3.1 general (healthy) public
   3.2 outdoor active (healthy) public
   3.3 patient
      3.3.1 respiratory disease patient
      3.3.2 heart patient

Figure 2: Default user profile typology in MARQUIS

substance); respiratory disease specialists are assigned the feature 'very sensitive in general' because, e.g., asthma patients suffer from elevated concentrations of any pollutant substance in the air. As far as the general public is concerned, it can be 'not particularly sensitive' (general public), 'sensitive' (people doing outdoor sports and thus more exposed to air pollution), 'very sensitive to ozone' (heart patients), or 'very sensitive in general' (respiratory disease patients).[5]

Each default profile contains for each communication channel (printed media, TV, mobile phone (SMS, MMS), and internet/WAP) a setting of content selection parameters that is considered adequate for users with this profile.[6] Thus, the domain professional is scheduled to

[5]Note that our user profile typology is still rather crude; for instance, it does not incorporate children, who are particularly sensitive to air pollution. As a matter of fact, medical studies show that each individual reacts differently with respect to air pollution (Oglesby *et al.*, 2000). In other words, ideally, an AQ information service would be personalized to such a degree that it would deal with personalized threshold concentrations for each pollutant substance. However, this would presuppose an extensive medical check of each individual subscribing to the service, which is not feasible. For user modelling, this would mean that profiles would have to be derived from personal records of the individuals—similar to (Cawsey *et al.*, 2000).

[6]Only selected user profiles have been fully realized in the operational version of MARQUIS. Some branches

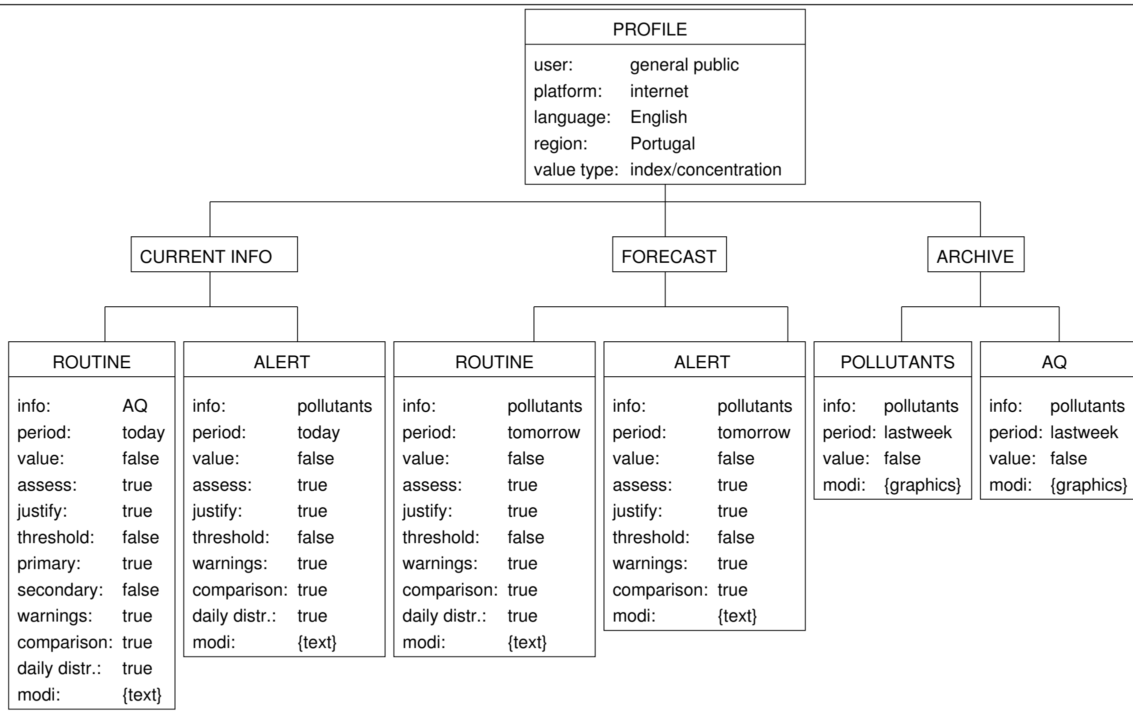


Figure 3: Default content selection parameter setting for general public, internet

receive pollutant concentrations listed in tables without further explanatory information.

A user subscribing to the MARQUIS service chooses a default profile that suits them best for a given communication channel. Cf. the setting for the profile *general public, internet* in Figure 3.

3.2 Personalization of User Profiles

Once registered, the user can personalize the profile he/she has chosen during the registration procedure in order to make it fit better their needs.[7] Personalization is possible along the lines

of the typology have been reworked in the meantime to incorporate changes that became necessary after the evaluation; we show the current version of the typology. For the original typology, see (Wanner *et al.*, 2007).

[7]Note that we distinguish between individuals to receive AQ information and information brokers, i.e., the press, radio channels, TV channels, web portals, etc. Information brokers often provide their own strictly defined

shown in Table 4.

Some attributes are predetermined by the type of the user and cannot be customized. For instance, 'value=false', 'judgement=true' for general public requests a rating but no index/or concentration; if an alert threshold is reached or surpassed, the associated health warning is shown regardless of whether this option was switched off by the user or not. In order to modify the fixed attributes, the user has to re-register under the corresponding user profile.

4 Overview of the MARQUIS Architecture

As illustrated in Figure 4, MARQUIS has a "two-pipe" architecture.

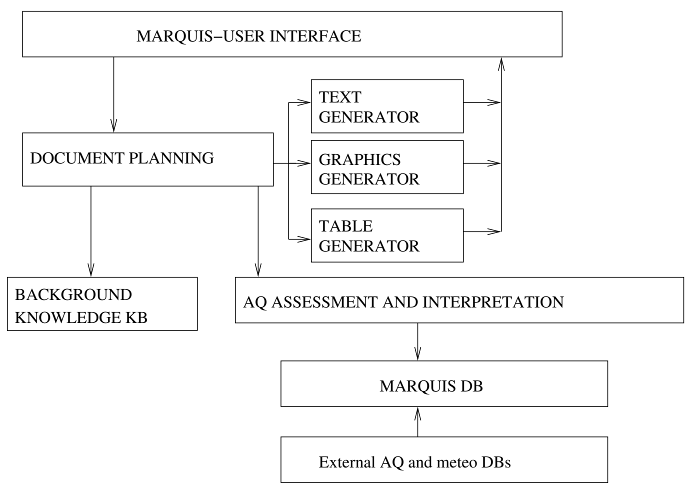


Figure 4: Architecture of the MARQUIS service

In the first pipe, the periodically measured and locally stored concentrations of air pollutant substances are delivered to the central MARQUIS DB, converted into the MARQUIS proprietary format and assessed and interpreted by the *Air Quality Assessment and Interpretation Module* (AQAIM). AQAIM is divided into two major submodules. The first submodule

profiles which deviate considerably from the profiles discussed here.

Table 4: Possible personalization of the default user profiles

time covered by the bulletin	☐ recent (current), ☐ forecasted (future) and/or ______ archive (past)[1] [1](specify the day, week, or the number of days in the past (within the span of a month) for which the information is requested)
air quality info*	☐ air quality index ☐ air quality index + mention of primary pollutants ☐ air quality index + mention of primary and secondary pollutants
pollutant concentration info**	______ list of pollutant substances[2] [2](list the pollutants for which information is requested)
alerts	☐ no alerts ☐ legal threshold alerts O_3 ___ PM10 ___ CO_2 ___ NO_X ___ SO_2 ___ AQI ___ personalized alerts[3] [3](specify the pollutant concentrations /indexes whose surpassing triggers an alert message)
delivery mode	___ o'clock[4] O_3 ___ PM10 ___ CO_2 ___ NO_X ___ SO_2 ___ AQI ___ concentrations/indices[4] ☐ will be retrieved by the user [4](specify the time respectively concentrations/indices)
language	___ language[5] [5](one of the eight MARQUIS languages)
presentation mode for * and **	☐text ☐table ☐graphics
region	______ name of a MARQUIS region
location	______ name of a monitoring station in the selected region

deals with the interpretation of the AQ numeric time series with respect to their relevance to the addressees; the second deals with the geographical pollutant extrapolation. Given that it is the first submodule which is critical for the evaluation of the MARQUIS service as an AI service, we focus in this article on it.

In the second pipe, the *Document Planning Module* (DPM) is triggered by a user information request. A request may come from the user profile demon, which continuously surveys whether the conditions to solicit information for any of the registered users are fulfilled, or from the user themselves.

Once a request is detected by the MARQUIS user interface, the document planner receives from the server the profile of the user in question and the AQAIM output structure for the time period in question (also determined by the user).

The document planner selects the corresponding knowledge from the AQAIM output structure and the background KB and produces a text plan, assigning to the sentence generator, the table generator and the graphic generator the text plan fragments to be realized in the corresponding mode. After the chunks of information are generated by the corresponding generators, they are merged together into a single document and delivered to the user via the MARQUIS client interface of the corresponding communication channel. Figure **??** displays, for illustration, a sample bulletin in English, French and Spanish generated by MARQUIS for the web-based service, targeted at general public with no personalization

5 Interpretation of AQ Numeric Time Series: Turning Data into Information

To be useful for an untrained user, the numeric time series must be assessed and the relevant information must be distilled from them. Furthermore, possible correlations between the different time series as well as between values of the time series and background information must

```
Freiburg-Mitte, 25 February '07 01:44
The air quality index is 3, which means that the air quality is satisfactory.
This is due to the ozone concentration.  The NO2 concentration, the SO2
concentration and the PM10 concentration do not have influence on the air quality.
The current air quality index (3) is the highest today.  The lowest was 2 (at
midnight).  Between midnight and 7AM, the air quality index remained stable at 2
and between 8 AM and 9PM, it remained stable at 3.
...

L'indice de qualité de l'air est de 3, ce qui signifie que la qualité de l'air
est satisfaisante.  Cela est dû à la concentration d'ozone.  La concentration
de dioxyde d'azote et la concentration de dioxyde de soufre ne contribuent pas à
l'indice.  L'indice de qualité de l'air actuel (3) est le plus élevé.  Le plus bas
était de 2 (à minuit).  Entre minuit et 7h, l'indice de qualité de l'air est resté
stable à 2 et entre 8h et 9h, il est demeuré stable à 3.
...

El índice de calidad del aire es 3, lo que significa que la calidad del aire es
aceptable.  Esto es debido a la concentración de ozono.  La concentración de
dióxido de nitrógeno y la concentracin de dióxido de sulfuro no contribuyen al
índice.  El índice de calidad del aire actual (3) es el máximo.  El mínimo era de
2 (a medianoche).  Entre medianoche y las 7:00, el índice de calidad del aire se
mantuvo estable en 2 y entre las 8:00 y las 9:00, permaneció estable en 3.
```

Figure 5: Fragment of a sample bulletin in English, French and Spanish

be determined.[8]

The four major time series that AQAIM processes are: the time series of observed and forecasted concentrations of the main pollutant substances and the time series of observed and forecasted meteorological conditions. The pollutant and AQ indices are calculated during the interpretation of the pollutant concentration time series. As mentioned above, the meteorological time series are used as auxiliary series for the interpretation/assessment and forecasting.

Calculation of the European and regional AQIs: The AQI is based on the highest concentrations among the air pollutant substances—the *primary pollutants*. Substances that do not contribute to the index due to their low concentrations are considered *secondary pollutants*. The scale of the AQI as well as the calculation procedure is country- or region-specific. For instance, in Germany it ranges from 1 to 6, in France from 1 to 8, and in Catalonia from -100 to $+100$.

The quantitative scale of any AQI is mapped onto a regional qualitative scale. For instance, the qualitative scale of the Finnish AQI is [good, satisfactory, fair, poor, very poor]. Along with the AQI, individual pollutant indices with a comparable quantitative and qualitative scale are common.

All indices are computed for each region. That is, the Finnish index is computed not only for FIN, but also for the other four MARQUIS regions; the same applies to the BW, CAT, PT and US indices—such that users at home in a given region can consult the AQ in other regions with their "home index". This supports the cross-border view on AQ.

[8] Note, however, that AQAIM is not user-tailored in the sense that it assesses the time series from the perspective of the addressee who might have requested AQ information. Rather, AQAIM assesses a given time series with respect to its relevance for each and all types of addressees who might be interested in AQ information. It is the discourse planning module that selects the content that is of relevance to the addressee for whom the information is being generated.

Table 5: Concentration–health impact association for Finland

Index	Color	Class	Health impact	Other long-term impacts
151–	violet	very poor	Adverse effects possible on sensitive subpopulation	Clear impacts on vegetation, material impacts
101–150	red	poor	Adverse effects possible on sensitive individuals	Clear impacts on vegetation, material impacts
76–100	orange	fair	unlikely effects	Clear impacts on vegetation, material impacts
51–75	yellow	satisfactory	very unlikely effects	mild environmental impacts
0–50	green	good	no health effects	mild environmental impacts

Assessment of Health Relevance of Pollutant Concentrations: For each region, the health relevance of given pollutant concentrations or index intervals is determined by matching the measured or forecasted concentrations or indexes with manually compiled concentration–health impact tables. Table 5 displays the table for Finland.

In addition, regional, national and WHO-authored precautionary measure advices related to specific index intervals are considered during information generation.

Significant Changes of Pollutant Concentrations: A numeric time series can be plotted as a function. Thus, it is not surprising that the topics of interest during the interpretation and assessment of an air quality time series are very much similar to those addressed in mathematical curve analysis;[9] due to their general nature, we expect these topics of interest to be relevant to any numeric time series, cf. Figure 6:

1. the starting and end values of the interval over which the series is defined,
2. the significant relative and absolute minima and maxima of the series,
3. significant positive and negative gradients of the series,
4. time stamps for values within the series that are higher than legal or user-defined thresholds.

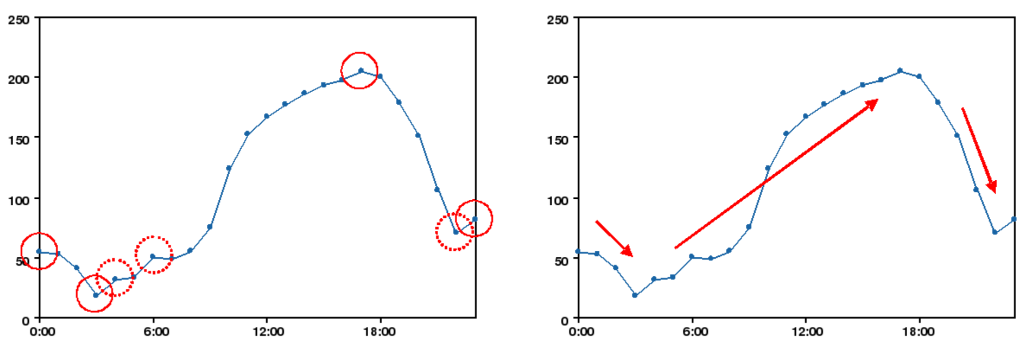


Figure 6: Curve analysis of the ozone time series at the monitoring station Ludwigsburg, BW, 2005-05-28

Air Quality–Meteorological Condition Correlations: To provide motivation (or justification) for the distribution of the concentration of a pollutant, correlations between the meteorological time series episodes and the relevant features of the pollutant/AQI time series (cf. 1.–4. above) are determined. Depending on the region, this is done by multiple regression,

[9]The case of the time series is somewhat more complex than a standard curve analysis because over some parameters a significance, i.e., interpretation/assessment, function is defined.

machine learning, or expert rule models. Table 6 displays some AQI–weather condition associations in the case of Portugal. For justification, the conjunction of the weather conditions that apply is taken.

Output Structure of the Interpretation Module: The output structure provided by AQAIM is divided into three main sections: a. Current with the data for the current day, b. Forecast with the data for the next day, and c. Archive with the data for a predefined number of previous days. Each section may include information about the individual polluntant concentrations, the pollutant index, and the global AQI.

The individual pollutant and AQI information consists of the following five information elements:[10]

Time-concentration tuples. The concentration time series of the pollutant in question for a given day.

Exceedance sequence. How often and how long a threshold was/will be surpassed, what type of threshold it is (e.g., information or alarm threshold) and what was/will be the (absolute and relative) maximal exceedance.

VIP sequence. The most prominent values in the concentration time series of the pollutant—for instance, the starting value, the last available value, the minimum, maximum and average.

Delta sequence. The differences between specific value pairs in the concentration time series (such as minimum and maximum, first and current, etc.).

VIC sequence. The most prominent changes in the concentration time series, which are specified by three attributes: GRADIENT (*sharp* or *slight*), TENDENCY (*raise* or *drop*), and

[10]'VIP' stands for "Very Important Point" in the a pollutant concentration / AQI time series over a given day, 'VIC' "Very Important Change" in a pollutant concentration / AQI time series.

Table 6: AQI–weather condition correlations for Portugal ('X' is the radiation strength constant)

Index	Weather conditions likely to correlate with the following conditions
bad	– anticyclone with weak wind – extended stability – depression from the north of Africa with a chain of SE in the continent carrying dust from the desert – if ozone primary: (OR radiation $>$ X, continuous hot weather)
weak	– anticyclone with weak wind – if ozone primary: (OR radiation $>$ X, continuous hot weather) – depression from the north of Africa with a chain of SE in the continent carrying dust from the desert – extended stability
medium	diverse meteorological situations with pleasant conditions
good	– front with moderate activity – moderate wind
very good	– moderate to strong wind – low temperatures – precipitation – fronts with moderate activity

STEADINESS (*true* or *false*).

Consider for illustration Figure 7, which shows a fragment of the XML output structure provided by AQAIM.

Within some regions, zones that correspond to the catchment areas of regional air quality information broker clients are defined. For each zone, AQAIM provides, in addition to the information illustrated in Figure 7, an overall assessment of the zone.

6 Generation of AQ Summaries

Starting from the AQAIM output structure, the background KB and the user profile settings, MARQUIS generates user-tailored reports in the requested language. In what follows, we describe how the two major tasks implied by report generation, discourse planning and linguistic realization, have been addressed.

6.1 Planning the Discourse

As is common in text generation, the discourse planning module in MARQUIS is divided into two parts: content selection and discourse structure determination. Each task and subtask of the planning module is implemented in XSLT as a specialized set of operators (conditional blocks or *templates* in the XSLT terminology), which are called in a pipeline with their corresponding inputs and outputs by a Java program. XSLT is a powerful language that can be used for transforming one or more XML inputs into an XML output; therefore it is particularly suited for the task at hand. XSLT has already been used successfully in NLG in general (Wilcock, 2001) and in text planning in particular (Foster & White, 2004).

```
−<assessment>
  −<site site_name="Karlsruhe-Nordwest" site_location="KA-Nordwest">
    −<day date="2003-06-12" ref_date="2003-06-11">
      −<aq>
        <index tendency="1" value="5" region="BW/>
          −<time_concentration_sequence regions="BW US">
              ...
              <t_c_tuple received="17:00" conc="5" poll="O3"/>
          </time_concentration_sequence>
          −<vip_sequence regions="BW US">
              <vip type="abs_min" st="02:00" et="08:00" value="2"
                    poll="O3 NO2"/>
              ...
          </vip_sequence>
          − <vip_delta_sequence regions="BW US">
               <vip_delta vip1="abs_max" vip2="abs_min" value="3"/>
               ...
          </vip_delta_sequence>
          −<vic_sequence regions="BW US">
              <vic sv="2" ev="2" grad="constant" tendency="constant"
                    steady="true" st="02:00" et="08:00"/>
              ...
          </vic_sequence>
      </aq>
      −<pollutants>
        −<pollutant name="O3" scale="µg/m^3">
          −<indices>
            ...
</assessment>
```

Figure 7: Fragment of the output structure of the AQAIM

6.1.1 Content Selection

For reasons of optimized implementation, the content selection task is divided into a *fact production and aggregation* substask, which groups elements of the input structure that together form discourse structure spans, and a *fact selection* subtask, which extracts the spans that are relevant to the user.

Fact Production and Aggregation. Fact production is dealt with in several iterations. First, elements of the AQAIM output are selected according to the region specified in the user profile (e.g., select the AQ index for Finland since the user is from Finland). Next, further background information (such as ratings of the indices encountered in the AQAIM output, health-related advices, time intervals, etc.; see Section 2.4) is added. For each of the elementary elements thus introduced, a unique reference is generated that is used subsequently by the linguistic generator to produce textual references between spans. The elementary elements are then rearranged (or aggregated) into more complex structures that are used as information units in the final text plan such as the set of all pollutants, and the set of all primary pollutants; cf. Figure 8 for illustration, where primary pollutants are aggregated together.

As in some other report generators (see, e.g., Portet *et al.* 2009) in MARQUIS, aggregation is thus done at the content level. Linguistic aggregation as described, e.g., in (Dalianis, 1999) proves not to be essential. This is mainly because: (i) access is available to the content structures of the entire text plan, (ii) the content structures are detailed enough to control all repetitions, (iii) the semantemes within the semantic structures are not decomposed into their meaning parts such that no additional repetition is introduced during linguistic generation.

In the last iteration of fact production, semantic relations between the items of the elementary spans are introduced. For example:

- Introduce a CAUSE relation between the air quality index/concentration/rating and the meteorological condition.

```
<xsl:template match="pollutants" mode="primary">
<xsl:if test=".//pollutant[@role='primary']">
<primary>
<xsl:for-each select="pollutant[@role='primary']">
<pollutant>
<xsl:copy-of select="@* [name() != 'role']"/>
<xsl:copy-of select="*"/>
</pollutant>
</xsl:for-each>
</primary>
</xsl:if>
</xsl:template>
```

Figure 8: A sample fact aggregation XSLT template

- Introduce an IMPLICATION relation between the air quality exceedence/index/concentration/rating and the health risk message.
- Introduce an EQUIVALENCE relation between the air quality index/concentration and the air quality rating.
- Introduce a CONSTITUENT relation between the argument of a function defined over a time-series (i.e., primary pollutants) and a value of this function (i.e., air quality index).

Fact Selection. The fact production task can be considered as a "preparatory" task that makes the first region-oriented content choices and transforms fragments of the AQAIM output structure into a planner processible format. In contrast, the fact selection task is largely user model driven. It consists in the application of the content selection operator XSLT templates. For instance, an XSLT template is defined for the selection of current information. It reads as follows:[11]

> If the user profile required current information and current information is available, then include current information; else if the user profile required current information, but no current information is available, then include a *no data* message with the relevant reference.

Further content selection operator XSLT templates include:

- SELECT the pollutant and its concentration if the concentration is above an alarm threshold (regardless of whether selected by user or not).
- SELECT health warnings associated to a pollutant if requested in the user profile or if its concentration reached or is above an alarm threshold.
- SELECT VIPs, VICs and time stamps in which the concentration of a pollutant is above a threshold concentration if the user requires the daily distribution of a pollutant substance.
- SELECT VICs with a sharp gradient and a non-constant tendency. If none are available, select a VIC in case it lasted longer than a set constant: `<vics hours_duration="3"/>`.
- SELECT time-pollutant concentration tuples if at least two different concentrations have been measured, and if the user opted in their profile personalization for graphics or tables as communication mode.
- SELECT secondary pollutant concentrations if required by the user in their user profile settings.

The result of fact selection is a sequence of isolated elementary information spans, with semantic relations defined between the units of each span. This sequence is to be casted into a discourse structure and verbalized by the subsequent linguistic realization module.

[11]Due to the lack of space, we do not cite here actual XSLT code.

6.1.2 Discourse Structure Determination

Analogously to Rösner and Stede (1992), in the operational version of MARQUIS, the discourse structure is defined in terms of a text schema (McKeown, 1985) in which between the elements of the schema discourse relations in the sense of the Rhetorical Structure Theory, RST (Mann & Thompson, 1987) are established.[12]

Figure 9 shows a fragment of an output text plan in XML format. It consists of the elementary rhetorical tree that links by a NEGATIVE JUSTIFICATION relation the AQ rating to its secondary pollutants' (ozone, sulfur dioxide and carbon monoxide) concentrations. This elementary rhetorical tree is wrapped in an ELABORATION relation of type "topic-focus" that specifies the current topic—which is the AQ secondary pollutants in the current information section. This information helps in the ordering task.

Text Schema in MARQUIS. The basic text schema that has been worked out in collaboration with AQ information experts for use in MARQUIS is shown in Figure 10.

Mapping semantic relations onto discourse relations. In the process of the creation of the document plan, semantic relations identified between units of the AQAIM output structure are mapped onto RST relations.[13] MARQUIS uses a restricted set of seven RST relations that proved to be of relevance in the air quality domain: JUSTIFICATION, CONSEQUENCE, EVALUATION, EVIDENCE, INTERPRETATION, CONTRAST, and LIST. ANALOGY, is subdivided into three more specific relations (EQ(ual)-ANALOGY, POS(itive)-ANALOGY and NEG(ative)-ANALOGY). Table 7 displays the semantic-discourse relation mappings. 'EQ' stands for 'EQUAL'. 'COORD' is inspired by WordNet's COORDINATE relation and means 'of the same type'.

[12]Currently, we are working on a more flexible discourse structure derivation in which, as, e.g., in BT-45 (Portet *et al.*, 2009), the schema is substituted by a dynamic ordering of spans and construction of a discourse graph.

[13]The identification of the semantic relations is rather straightforward in that it uses domain-specific manually crafted rules.

```
<span id=”N65932” relation=”topic-focus-elaboration” modes=”text”>
  <span id=”N65938” node=”nucleus”>
    <infounit id=”N65941”>
      <topic id=”N65943” section=”current” type=”aq”
        subtypes=”secondary” alert=”false”/>
    </infounit>
  </span>
  <span id=”N65951” relation=”negative-justification” node=”satellite”>
  <span id=”N65956” node=”nucleus”>
    <infounit id=”N65959” order=”1.2”>
    <rating id=”N65961” ref=”S65568” label=”1/6” possessor=”aq”
      region=”Catalunya”/>
    </infounit>
  </span>
  <span id=”N65973” relation=”list” node=”satellite”>
    <span id=”N65977” node=”nucleus”>
      <infounit id=”N65980” order=”1.4.1”>
      <concentration id=”N65982” possessor=”O3”/>
    </infounit>
  </span>
  . . .
</span>
```

Figure 9: Fragment of an ordered text plan

Table 7: Semantic-discourse relation mappings (‘D’ = constant of a concentration, defined by the user; ‘T’ = concentration threshold; ‘VIC’ and ‘VIP’ as introduced in Footnote 11; for the last mapping: a VIC is relevant if the gradient is sharp and the tendency is not constant, or if the gradient and tendency are constant, but duration is above a constant)

semantic relation	discourse relation	conditions
CAUSE(arg_1,arg_2)	JUSTIFICATION(n,s)	arg_1=‘meteo’, arg_2=‘rating’; n= ‘rating’, s= ‘meteo’
IMPLY(arg_1,arg_2)	CONSEQUENCE(n,s)	arg_1=‘exceedance’|‘rating’|‘index’|‘concentr.’ arg_2=‘health-risk’ n=‘health-risk’, s=‘exceedance’|‘rating’|‘index’|‘concentr.’
EQ(arg_1,arg_2)	EVALUATION(n,s)	arg_1=‘index’|‘concentr.’, arg_2=‘exceedance’ section $\neq$ <ALERT>; n=‘index’|‘concentr.’, s=‘exceedance’
EQ(arg_1,arg_2)	EVIDENCE(n,s)	arg_1=‘index’|‘concentr.’, arg_2=‘exceedance’; section = <ALERT> n=‘exceedance’, s=‘index’|‘concentr.’
EQ(arg_1,arg_2)	INTERPRETATION(n,s)	arg_1=‘index’|‘concentr.’, arg_2=‘rating’ n=‘index’|‘concentr.’, s=‘rating’
COORD(arg_1,arg_2)	CONTRAST(n,s)	(arg_1 = current VIP, arg_2 = first VIP) OR (arg_1 = absolute min VIP, arg_2 = absolute max VIP) $\Delta(arg_1,arg_2) > D$; n=arg_1, s=arg_2
COORD(arg_1,arg_2)	LIST(n_1,n_2)	(arg_1 = current VIP, arg_2 = first VIP) OR (arg_1 = absolute min VIP, arg_2 = absolute max VIP) $\Delta(arg_1,arg_2) \leq D$; n_1=arg_1, n_2=arg_2
COORD(arg_1,arg_2)	EQ-ANALOGY(n,s)	(arg_1 = current VIP, arg_2 = absolute min. VIP) OR (arg_1 = current VIP, arg_2 = absolute max VIP) $\Delta(arg_1,arg_2) = 0$; n=arg_1, s=arg_2
COORD(arg_1,arg_2)	NEG-ANALOGY(n,s)	(arg_1 = current VIP, arg_2 = absolute min. VIP) OR (arg1 = current VIP, arg_2 = absolute max VIP) $\Delta(arg_1,arg_2) \leq D$, $arg_1 < arg_2$; n=arg_1, s=arg_2
COORD(arg_1,arg_2)	POS-ANALOGY(n,s)	(arg_1 = current VIP, arg_2 = absolute min. VIP) OR (arg_1 = current VIP, arg_2 = absolute max VIP) $\Delta(arg_1,arg_2) \leq D$, $arg_1 > arg_2$; n=arg_1, s=arg_2
COORD(arg_1,arg_2, ...)	LIST(n_1,n_2,...)	arg_i = VIP, $arg_i > T$; n_i=arg_i
COORD(arg_1,arg_2, ...)	LIST(n_1,n_2,...)	arg_i = relevant VIC; n_i=arg_i

1. AQ index and rating
2. Primary pollutants
3. Secondary pollutants
4. For each primary pollutant
    4.1 concentration or index
    4.2 rating
    4.3 VIPs and VICs
    4.4 alert (if applicable)
    4.5 health risks
5. Archive information
    5.1 pollutant concentrations/indices over Δ days
6. Forecast for each pollutant selected by the user
    6.1 expected concentration/index
    6.2 justification
    6.3 alert (if applicable)

Figure 10: The text schema used in MARQUIS

As the table shows, EQ and COORD can correspond to several discourse relations. Which one is to be chosen in a given case depends on the concrete context conditions (specified in the third column of the table). For instance, EQ, which holds between the concept configuration that expresses the concentration or index of a substance and the configuration that codifies the exceedance of a threshold by this concentration/index, is mapped onto EVIDENCE if the information is to be located in the ALERT section and onto EVALUATION otherwise ('n' stands for "nucleus" and 's' for "satellite").

6.2 Linguistic Realization

The theory underlying the linguistic realization in MARQUIS is the Meaning-Text Theory (MTT) (Mel'čuk, 1988). MTT has traditionally been popular in text generation due to its dependency-based multistratal linguistic model, which allows the developer, on the one hand, to select for each input structure a degree of abstraction that suits best the application in question, and, on the other hand, to keep the generation resources modular and simple (Iordanskaja *et al.*, 1991; Iordanskaja, 1992; Goldberg *et al.*, 1994; Caldwell & Korelsky, 1994; Coch, 1998; Bohnet *et al.*, 2001).

In MTT, linguistic realization can be viewed as a sequence of transductions between structures of adjacent strata, starting from the stratum of the input structure. In MARQUIS, this is the *conceptual* stratum, which contains conceptual graph structures in the sense of Sowa (2000) derived from the document plan.

For each pair of adjacent strata $\mathfrak{S}_i$ and $\mathfrak{S}_{i+1}$, a separate rule-based grammar module $\mathfrak{G}^i_{i+1}$ is defined such that any well-formed structure S_{i_j} of $\mathfrak{S}_i$ can be mapped by $\mathfrak{G}^i_{i+1}$ onto a well-formed structure S_{i+1_k} of $\mathfrak{S}_{i+1}$, with S_{i_j} and S_{i+1_k} being equivalent with respect to their meaning. As a rule, the mapping requires access to information concerning the units of S_{i_j} and S_{i+1_k}, which is stored in dictionaries. The grammar modules are implemented in MATE (Bohnet *et al.*, 2000; Bohnet, 2006), with the individual rules having the following format; see (Bohnet, 2006:39ff) for details:

leftside (ls): $<g_i>$
rightside (rs): $<g_{i+1}>$
rightcontext (rc): $<g'_{i+1}>$
conditions (cd): $<\text{Boolean expr. over } \mathfrak{D}_{conc} \cup \mathfrak{D}_{sem} \cup \mathfrak{D}_{lex} \cup \mathfrak{S}_i \cup \mathfrak{S}_{i+1}>$
correspondences (cr): $\{n_{i_j} \Leftrightarrow n_{i+1_k}\}$

with g_i being a graph defined over the node and arc alphabets of $\mathfrak{S}_i$, g_{i+1}, and g'_{i+1} being defined over the node and arc alphabets of $\mathfrak{S}_{i+1}$; $\mathfrak{D}_{conc}$, $\mathfrak{D}_{sem}$, $\mathfrak{D}_{lex}$ being the conceptual, semantic and lexical dictionaries; and $n_{i_j} \in g_i$, $n_{i+1_k} \in g_{i+1}$. The application of a rule consists in the identification of an isomorphic image of g_i in a given source structure S_{i_j} and subsequent introduction of an isomorphic image of g_{i+1} in the target structure S_{i+1_k}, which is under construction. The statement '$n_{i_j} \Leftrightarrow n_{i+1_k}$' establishes a link between corresponding nodes in S_{i_j} and S_{i+1_k} in order to ensure that (i) information can be propagated from node to node across strata, and (ii) the isolated fragments of the target structure as introduced by the individual rules can be unified to a connected well-formed structure. A rule is applicable if the specified conditions are fulfilled. As indicated, conditions may be defined over all dictionaries and both strata.

The rules in $\mathfrak{G}^i_{i+1}$ are minimal in the sense that the left-hand side of each rule is maximally elementary from the linguistic perspective. Consider the following rule for illustration:

Rule 1 (Sample $\mathfrak{G}^{Con}_{Sem}$ rule)

```
ls:  ?Xcon{PTIM->?T{con="tomorrow"}}
rc:  ?Xsem{tense=FUT}
cr:  ?Xcon ⇔ ?Xsem
```

Rule 1 maps the conceptual time relation between the concept denoted by the variable '?Xcon' and the concept TOMORROW onto the tense feature "FUT" of the semanteme denoted by the variable '?Xsem'. '?Xsem' is specified in the right context slot ("`rc:`"), which means that the corresponding semanteme is assumed to have been already introduced into the target structure by another rule. Note that this rule is valid in the grammar of any language that has future tense since the only label, "tomorrow", is a (language-independent) concept label.

For the sake of efficiency, the definition of rules shared by several grammar modules has been given high priority in MARQUIS. As a consequence, the percenage of language-specific rules is, in general, quite low (between 8 and 20%, depending on the language and level of linguistic realization). Only for morphological realization, this percentage increases considerably (e.g., to about 59% for French and Polish). The topic of multilingual linguistic realization and efficient organization of grammar modules in MARQUIS has been addressed in detail in (Lareau & Wanner, 2007). In what follows, we use therefore for illustration mainly English.

Figure 11 illustrates the steps of linguistic realization in MARQUIS. The righthand side of the diagram displays the strata (and thus the types of linguistic structures) that come into play and the tasks assumed by the corresponding grammar modules. The lefthand side of the diagram shows the structures at the different strata during the generation of the sentence *This means that air quality is very poor*.

Starting from the conceptual structures as input, the linguistic realization module has to account for a number of sentence planning (or microplannning) tasks, which concern: 1. communicative (= information) structure determination, 2. syntactic structure determination, 3. lexicalization, and 4. referring expression generation. As mentioned in Subsection 6.1.1, aggregation is restricted to conceptual aggregation and is done in the document planning module. The task of clause and sentence chunking is reduced to the minimum in that each discourse unit identified in the document plan is by default realized as a separate sentence.

The tasks 1–4 are implemented in terms of rules distributed among different grammar modules as a pipeline of transductions.[14] Let us briefly sketch in what they consist.

Determination of the communicative structure. The communicative structure (CommS) of a sentence is constructed during the transition between the conceptual and semantic representations.[15] It is essential for the determination of the syntactic structures and lexicalization (see below). In MTT, CommS consists of eight communicative dimensions, each of which is defined in terms of a number of parameters (Mel'čuk, 2001). From these eight dimensions, MARQUIS uses three:

[14]As argued by Polguère (1998), this is a simplification since several of the above tasks, including, e.g., lexicalization, may require some backtracking and "look ahead". However, in operational report generation, this simplification is justified; see also (Reiter, 1994) with respect to pipeline architectures in NLG.

[15]Communicative structure is also often referred to in the literature as *information structure* (Lambrecht, 1994; Vallduví, 1995).

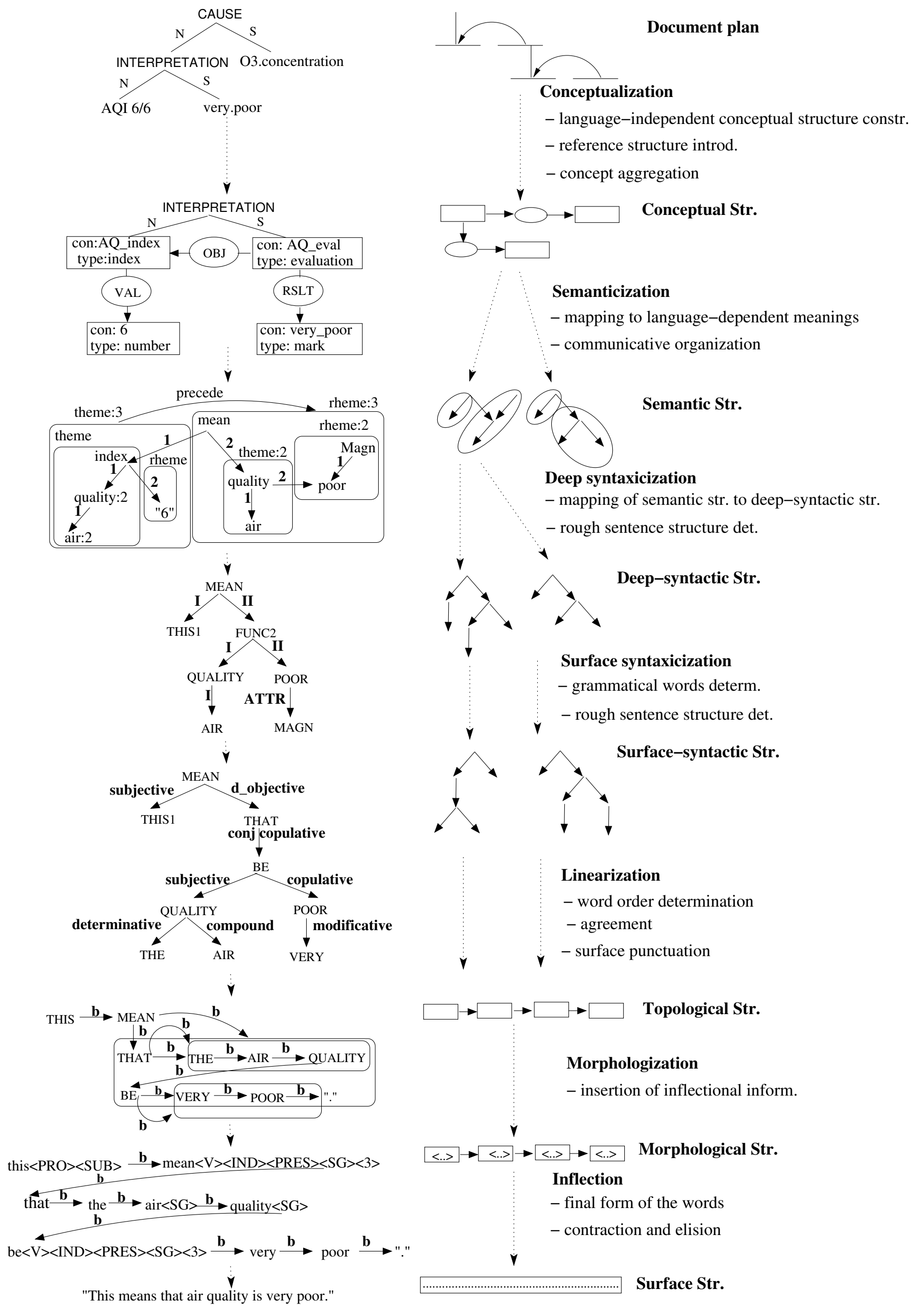


Figure 11: The steps of the linguistic realization process in MARQUIS

1. *thematicity* (with the parameters rheme/theme/specifier) partitions the semantic graph into communicative subgraphs identifying the theme (the topic of the sentence), the rheme (what is said about the topic, i.e., *comment*), and specifiers (which provide auxiliary information from a communicative point of view, such as time and location).

2. *givenness* (with the paramters given/new) identifies parts of the graph as 'new' respectively 'given' information.

3. *perspective* (with the paramers foregrounded/backgrounded/neutral) is used to mark subgraphs as being of particular importance / minor importance / neutral with respect to its relevance in the given context.

Thematicity is partially predetermined by *domain communication knowledge restrictions* (Kittredge *et al.*, 1991), and partially derived from the discourse structure. Domain communication knowledge restrictions include, for example, the fact that names of pollutants or parameters such as *air quality index* (but not its value) are by default thematic, since they are considered as the general topic of the text. Examples of the discourse structure-based derivation rules are: Elaboration(X,Y) $\Longrightarrow$ theme(X), rheme(Y) and Interpretation(X,Y) $\Longrightarrow$ theme(X+Y).

In each thematic and rhematic area, one node is marked as the *dominant node*. The dominant node of a communicative subgraph is the most salient of its area. For example, in the expression (*this is due to*) *the concentration of ozone*, the semanteme 'concentration' communicatively dominates 'ozone' (the concentration of ozone is a concentration), while in (*we found*) *ozone in a concentration of 30* $\mu g/m^3$, it is the opposite (ozone in a concentration of 30 µg/m^3 is ozone). The dominant node of a subgraph must be linked with the dominant node of another subgraph by predicate-argument relations, the typical scenario in general speech being that the dominant node of the theme is a semantic actant of the dominant node of the rheme.

Giveness information is already specified in the text plan and is thus mapped directly onto the communicative structure.

Perspective is used in MARQUIS mainly to resolve potentially conflicting communicative configurations when a sentence has two rhemes. For instance, in the sentence *The* PM_{10} *concentration (13 μg/m*^{3}*) is relatively low*, the value of the concentration and its qualitative evaluation are both rhematic but do not belong to the same rheme. The sentence, in fact, expresses two messages that we want to communicate: 1. *the* PM_{10} *concentration is 13 μg/m*^{3} and 2. *the* PM_{10} *concentration is relatively low*. But a coordination would be inappropriate here. So, in order to avoid getting a two-headed sentence, one rheme is marked as backgrounded, which is linguistically implemented in this context as a parenthetic construction.

Syntactic structure determination. The syntactic structure is largely determined by the CommS settings derived early in the process of sentence planning; the corresponding top-down algorithm goes back to Polguère (1990). First, the syntactic root is realized in that the communicatively most salient of the semantemes in the SemR is lexicalized. By default this is the dominant node of the rheme; in configurations where a parameter (such as 'concentration') is thematic and its value rhematic, it is the dominant node of the theme that is predicative and is realized as the syntactic root. Starting from the root, syntactic dependents are built, either as actants, if the corresponding semantemes are arguments of the predicate corresponding to the root, or as modifiers, if, on the contrary, it is the semanteme corresponding to the root that is an argument of the predicate expressed by the syntactic dependent. The procedure is repeated recursively until the whole tree has been built (and the semantic graph has been entirely mapped). In the case of coordinated propositions, each proposition is dealt with as if it were an independent one. After the trees for each are built up, their roots are connected via a coordination.

Lexicalization. As already pointed out by Kittredge & Polguère (2000), in domain-specific report generation, it is important to account not only for a flexible choice of individual words, but also for *idiosyncratic multiword expressions*. Apart from complex terms (such as *monitoring station*), which can be treated as individual words, these expressions include *collocations*, i.e., lexically restricted binary word cooccurrences of the type *concentration dropped*, *the wind blows*, *heavy traffic*, *strong wind*, *heavy pollution*, etc. MTT provides a theoretical means to encode and deal with collocations: the *lexical functions* (LFs) (Mel'čuk, 1996). LFs capture common semantico-syntactic patterns underlying various concrete lexical cooccurrence expressions. For instance, the expressions *heavy rain*, *high concentration* and *strong wind* are manifestations of a common underlying pattern. *Heavy* is to *rain* what *high* is to *concentration* or *strong* is to *wind*. In all cases, we have an adjective depending syntacticly on a noun and meaning, roughly, 'intense'. Although these adjectives are formally different, they can, in fact, be considered as different uses of the same *generalized lexeme* (Wanner, 1996a).

This generalized lexeme is identified by the label **Magn**, and its form is given in the dictionary as a function: Magn(*wind*) = *strong*, $_{\text{CAT.}}$Magn(*vent*) = *fort*, $_{\text{POL.}}$Magn(*wiatr*) = *silny*, Magn($increase_N$) = *significant*, $_{\text{FRE.}}$Magn(*augmentation*) = *considérable*, etc.[16] In total, about sixty of such generalized lexemes, or functions, have been defined. Their incorporation into the dictionary of a generator facilitates flexible and rich lexicalization and paraphrasing (Iordanskaja *et al.*, 1992; Wanner, 1996b); due to their language independence, they are especially useful in the context of multilingual generation.

To accomodate for the use of LFs, lexicalization in MARQUIS is done in two stages: *deep lexicalization* (DL) and *surface lexicalization* (SL). DL is done by the $\mathfrak{G}^{sem}_{dsynt}$ module. During DL, meaningful semantemes of a given SemS that can readily be expressed by a lexeme, i.e., open class lexemes that are not controlled by lexical cooccurrence, are mapped onto lexemes in the DSyntS, while meanings that correspond to recurrent patterns of lexical cooccurrence are

[16]See the sample lexical entry for DUE$_1$ in Figure 13, where the LF **Oper2** is used, and the discussion below.

cause { lex = cause1 //noun	causar { lex = causa //noun
lex = consequence //noun	lex = consequencia //noun
lex = cause2 //verb	lex = causar1 //verb
lex = responsible //adjective	lex = "a causa" //adverb
lex = because //adverb	lex = porque //adverb
lex = due1 //adjective	lex = debido1 //adjective
lex = due2 //adverb	lex = debido2 } //adverb
lex = therefore } //adverb}	

Figure 12: Sample entries of the semantic dictionaries of English and Spanish

mapped onto the corresponding LF names. The possible lexicalizations of a given semanteme are given in the *semantic dictionary*; cf. the entry for 'cause' and its Spanish equivalent, 'causar' in Figure 12. The lexemes expressing a given semanteme need not be of the same part of speech, or even have an identical diathesis (e.g., 'X causes Y' = 'Y is due to X'). Hence, conversives, for example, always appear in the same entry.

The choice of a specific lexeme depends on the syntactic context; for example, a nominal lexicalization will be chosen only if a noun is expected in the position in question.[17] Information on the part of speech and the *government pattern* (≈ alignment of the semantic and syntactic valency structures) of the candidate lexemes is given in the *lexical dictionary*. Consider, for illustration, the entries for CAUSE and DUE$_1$ in Figure 13.[18] The dictionary also contains the

[17] If several lexemes are appropriate in a given context, the choice is random.

[18] Arabic numbers identify semantic actants, while Roman numbers stand for deep syntactic actants. For example, the entry for DUE$_1$ reads like this: it is an adjective; its first semantic actant becomes its second syntactic actant, while its second semantic actant becomes its governor with the relation ATTR; its governor must be a noun; its second syntactic actant is realized in surface syntax as an adjective completive, it must be a noun and it is introduced by the preposition TO$_1$; it controls the collocations *be due* (support verb) and *partly due* (attenuation).

```
cause {dpos = V
        gp = {1 = I
              2 = II
              I = {dpos = N }
              II = {dpos = N }}}
```

```
due1 {dpos = Adj
      gp = {1 = II
            2 = ATTR
            ATTR = {dpos = N }
            II = {rel = adj_completive
                  dpos = N
           prep = to1}}
      Oper2 = be
      AntiMagn = partly}
```

Figure 13: Sample government patterns and LFs from the lexical dictionary

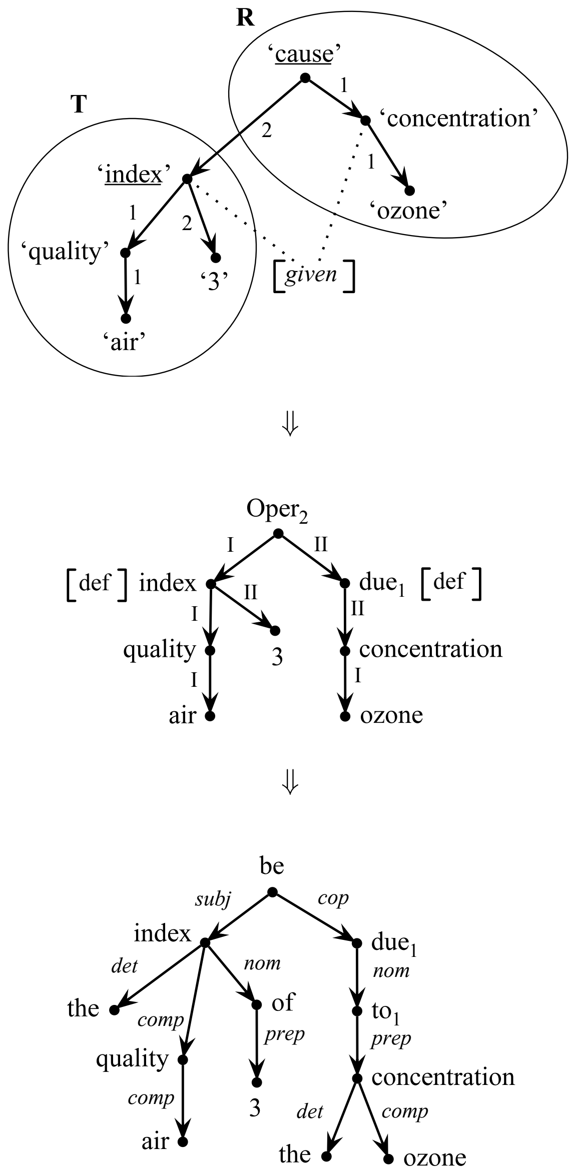
LFs associated with each of the lexemes. This is useful during DL, for instance, to resolve situations where a possible lexicalization does not fit into a given syntactic construction, as illustrated in Figure 14.

SL is carried out by the $\mathfrak{G}^{dsynt}_{ssynt}$ module. It consists in the substitution of LF names by the corresponding values and in the introduction of governed prepositions and grammatical words (articles, auxiliaries and so on).

The example in Figure 14 illustrates the mapping between the SemR, the DSyntS and the SSyntS for the sentence *The air quality index of 3 is due to the ozone concentration*; we leave aside the problem of tense.

Referring expression generation. MARQUIS accounts for the generation of three types of referring (or anaphoric) expressions: (i) introduction of definite/indefinite articles, (ii) pronominalization, (iii) sentential deictic anaphor generation, (iv) ellipses.

As mentioned above, the CommS indicates whether a given semanteme is new or given in-



- The dominant node of the rheme (‘cause’) is predicative; it is chosen as the entry node, and the root of the DSyntR is created.
- The semantic dictionary provides the verb CAUSE, but its diathesis in the lexical dictionary does not match the construction (a passive would be required here, but it is not appropriate for the domain).
- Oper2(*due1*)=*be* allows for the lexicalization of ‘cause’ as an adjective with a support verb.
- Some semantemes are marked as *given* because they were mentioned in a previous sentence. The corresponding deep syntactic nodes take the grammatical feature *def*.
- During the $\mathfrak{G}^{dsynt}_{ssynt}$ transition, grammatical words such as articles and prepositions are inserted.
- Surface lexicalization retrieves the value of Oper2(*due1*) from the lexical dictionary.

Figure 14: Mapping a SemR onto a DSyntR and a SSyntR

formation. In English, as well as in most other languages involved in MARQUIS, *given* markers are implemented on the surface by the definite article. The process is straightforward, except that it must take place in two steps. First, the $\mathfrak{G}^{sem}_{dsynt}$ module maps the *given* communicative marker onto a grammatical feature attached to the corresponding node of the DSyntR. Then, the $\mathfrak{G}^{dsynt}_{ssynt}$ module maps this grammatical feature onto a lexeme (the article *the* in the case of English, *der/die/das* in the case of German, etc.) in the SSyntS.

Pronominalization is somewhat more complex in that it interferes with regular lexicalization. To untangle this interference, pronominalization is done in two stages. First, it is determined which semantemes in the SemS are to be pronominalized. This is done by an auxiliary intermediate grammar module, $\mathfrak{G}^{sem}_{sem}$, which recopies the SemS and marks semantemes for pronominalization. As is the case for syntactic structure determination and lexicalization, it is the CommS that pilots pronominalization. Only the main node of a theme or rheme is a candidate for pronominalization. For example, if the theme of a sentence is identical to the theme of the previous sentence, it is marked for pronominalization. Marking a node triggers a number of rules that mark other nodes attached to it for deletion (to avoid having dependents on the pronoun). In order to verify the conditions for pronominalization, access to the previous sentence is required. So, each time a SemR is processed, it is kept in memory and added to the input for the next sentence. Consider, for illustration, the two SemRs in Figure 15. The first SemR would be realized as *The ozone concentration was high this morning.* The theme being repeated in the second sentence, its dominant node, ‘concentration’, will be marked for pronominalization, whereas ‘ozone’ will be marked for deletion, which will result in the sentence *It will be low this afternoon.*

Second, the actual introduction of pronouns takes place during the application of the $\mathfrak{G}^{sem}_{dsynt}$ module. The presence of a pronominalization mark on a node blocks the application of standard lexicalization rules and triggers a rule that maps the semanteme to the appropriate pronoun. This rule is language specific, since it introduces a specific lexeme (for example, IT in English). In German, Polish, etc., it needs to access the lexical dictionary to retrieve the gender of the

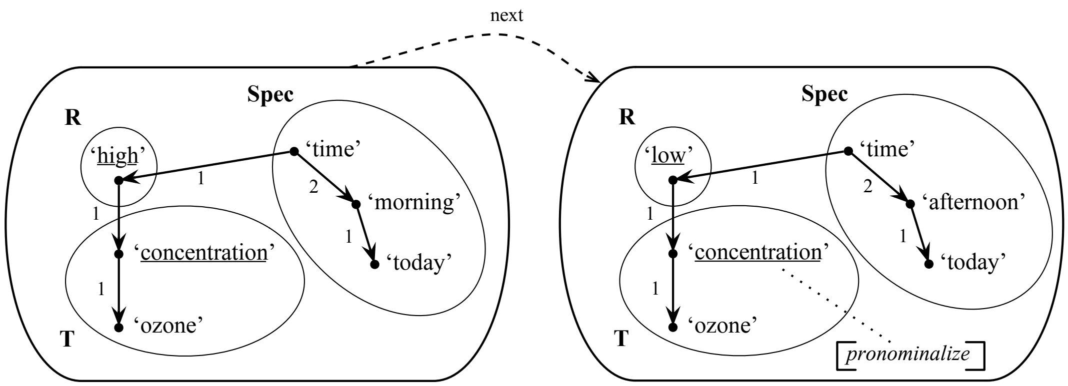


Figure 15: A repeated theme marked for pronominalization (T=theme, R=rheme, Spec=specifier)

noun being pronominalized.

Sentential deictic anaphors are also dealt with in the auxiliary $\mathfrak{S}_{sem}^{sem}$ module. Here again, we rely on the CommS to recognize the configurations that trigger this kind of anaphor. If the core semantic configuration of the previous sentence (i.e., its dominant nodes with their immediate dependents and modifiers) appear as the theme of a sentence, then it is replaced by a pronoun (THIS in English, CELA in French, DIES in German, ESO in Spanish, etc.). Thus, consider the two consecutive SemRs in Figure 16. The first corresponds to the sentence *The air quality index is 3.* The whole configuration appears as the theme of the next sentence; its dominant node will therefore be marked for sentence deictic anaphora, and the other nodes will be marked for deletion. The mark on the dominant node will block standard lexicalization and trigger a language-specific rule introducing the appropriate pronoun in the DSyntS. This will result in the sentence *This is due to the ozone concentration.*

Ellipses are to avoid useless repetitions of some words. In MARQUIS, they are mainly used in contexts such as *The ozone concentration is 30 $\mu g/m^3$. The highest ozone concentration was 100 $\mu g/m^3$*, where we do not want to repeat *ozone* in the second sentence. Nodes are marked

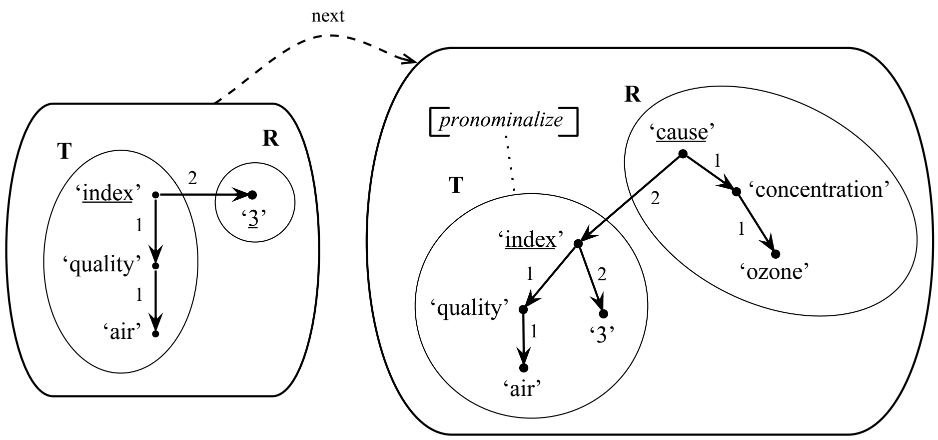


Figure 16: A sentence repeated as a theme marked for pronominalization

for ellipsis already in the conceptual structure. During text planning, a trace of entities already mentioned is kept and passed as attributes on the conceptual nodes. When, for example, a substance is marked as already mentioned, the corresponding conceptual node is realized as a semanteme but the semanteme is marked for ellipsis. This attribute will block the application of lexicalization rules in $\mathfrak{G}^{sem}_{dsynt}$. We do need the node at the semantic level however, in order to recognize patterns that trigger pronominalization and sentence deictic anaphora. Otherwise, the rules mentioned above would not apply, since the node 'concentration', in this example, would not have the same semantic dependents in the two sentences.

7 Evaluation

To assess the performance of the MARQUIS generator, two types of evaluations have been carried out. The first one was conducted periodically by the grammarians in the course of the development of the grammatical resources to identify failures and gaps and take remedial actons. This evaluation is described in detail in (Lareau & Wanner, 2007).

The second was conducted by users to examine the quality of the produced bulletins.

Table 8: Results of the evaluation

	% of the bulletins graded as G (from 1 = 'poor' to 5 = 'excellent')				
	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
comprehensibility	2.2	1.8	16.5	47.3	32.2
information order	0	0.4	26.8	30.4	42.4
fluency	0	5.3	48.7	38.0	8.0
grammaticality	0	2.2	10.3	31.7	55.8
word choice	0	1.8	9.4	50.0	38.8
content relevance	2.7	5.4	20.1	32.6	39.2
level of detail	0.4	6.3	31.7	40.2	21.4
missing content	0	2.2	14.3	17.9	65.6
appropriateness for decision support	7.6	7.1	15.3	19.6	50.4
general satisfaction	0	11.6	25.0	48.2	15.2

Bulletins in five languages were evaluated: Catalan, English, French, German, and Spanish. Finnish, Polish and Portuguese were not included in the evaluation due to the insufficient number of readers of these languages among the test users. In total, the evaluation pool contained 80 bulletins generated from varying AQ assessment plans; 16 in each of the five languages. Ten independent users were involved in the evaluations. Each of them evaluated between 16 and 64 bulletins generated for users with the default profile of general public according to the criteria in the questionnaire displayed in Figure 17.

The results of the evaluation are summarized in Table 8.[19] The table reflects the percentage of the ratings according to the ten criteria which fell into each of the five grade categories—from 1 (poor) to 5 (excellent).

[19]Due to the cross-lingual development of the grammatical and lexical resources (cf. Lareau & Wanner 2007), the content and language quality of the bulletins across languages were very similar, such that the quality figures below have been calculated across all evaluation sheets (rather than language-wise).

Text number:

Evaluator:

Please grade each criterion from **1** (poor) to **5** (excellent/high), if not stated otherwise.
Evaluate the quality of this text independently of other texts.

I. Language quality:

I.1 Comprehensibility of the language (   )
(Is the text understandable? Is it clear what the report communicates?)

I.2 The order in which the information is presented in the bulletin (   )
(To what extent does the order of the information coincide with your intuition of an adequate ordering of the presentation of the facts in an air quality bulletin?)

I.3 Fluency of the language in the bulletin (   )
(Does the text read well? Is it monotonic or repetitive?)

I.4 Grammaticality of the sentences in the bulletin (   )
(Is it correct Catalan/English/French/German/Spanish or are there grammatical mistakes?)

I.5 The appropriateness of the chosen words in the bulletin (   )

(Are the words well chosen or would you use other wordings?)

II. Adequacy of the content in the bulletin:

II.1 Relevance of the content communicated in the bulletin (   )
(Are the facts that the report mentions relevant?)

II.2 Level of detail of the information (   )
[if less than 4: too much detail (   ) not enough detail (   ); tick one]

II.3 Coverage of the relevant content in the bulletin (   )
(Are you missing any content in this report? Is there anything else you would have added?)

III. Informativeness of the bulletin

III.1 Appropriateness for decision support (   )
(Is the information provided in the report suitable to support people in their decisions?)

IV. Overall

IV. General satisfaction with the bulletin (   )
(How positive / negative is your overall impression of the report? Would you use it to check the air quality in your town or do you prefer other ways of air quality presentation?)

The quality of the bulletins is thus in general considered good. Let us however analyze in more detail the criteria according to which a noticeable percentage of bulletins received the negative grades 1 or 2. These are 'comprehensibility', 'fluency', 'content relevance' and 'level of detail'. The study of the bulletins that received low grades for 'comprehensibility' shows that the great majority of them contains a passage like the one shown in Figure 18, (a).

(a). `The air quality index is 6, which means that the air quality is very poor. This is due to the ozone concentration. The nitrogen dioxide concentration and the carbon monoxide concentration do not influence the index.`
`The ozone concentration (25 µg/m`^{3}`) is very low. As a consequence, no harmful effects on human health are expected.`

(b). `The PM10 concentration (17 µg/m`^{3}`) is low. The PM10 concentration is due to strong winds.`

(c). `The sulfur dioxide concentration is very low (70 µg/m`^{3}`). Thus, no harmful effects on human health are expected.`
`Between midnight and 3 AM the sulfur dioxide concentration remained stable at 1. Between 8AM and 10AM, the sulfur dioxide concentration decreased considerably from 6000 to 4000 and in the late morning, it decreased from 400 to 100. ...`

Figure 18: Fragments of English bulletins

This obvious incoherence results, on the one hand, from the way the AQI is calculated by the region-specific AQ model, namely as the average of the last eight hours, and, on the other hand, from the inference weakness of the AQAIM in MARQUIS: it does not infer that the ozone concentration fell drastically during the last 8 hours.

The problem with fluency is first of all due to a not sufficiently rich anaphora generation in specific constellations; e.g., Figure 18, (b) and parenthetical constructions that are often used to communicate the actual concentration of a pollutant.

As far as 'content relevance' and 'level of detail' are concerned, many users considered an

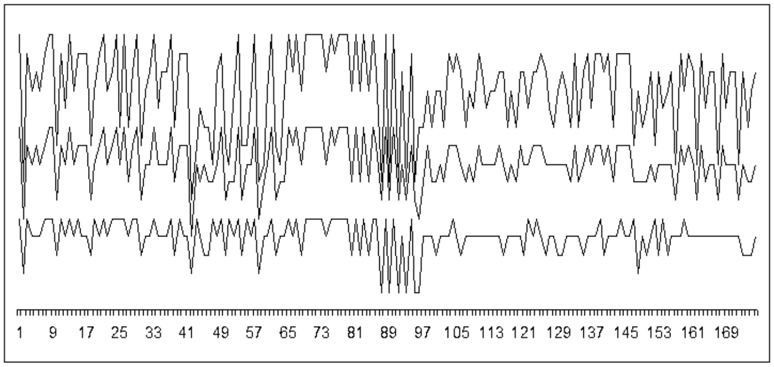


Figure 19: Correlation between criteria grades; upper curve: ‘appropriateness for decision support’ grades; middle curve: ‘content relevance’; lower curve: ‘comprehensibility’

interval-like presentation of the past concentrations of a primary pollutant substance irrelevant; cf. Figure 18, (c).

As the graphic in Figure 19 shows, there is a direct correlation between the grades of ‘comprehensibility’, ‘content relevance’ and ‘appropriateness for decision support’: if a bulletin was considered incomprehensible or its content was considered irrelevant, it was also marked as poor with respect to ‘appropriateness for decision support’, and vice versa. The general satisfaction of a user with a bulletin coincides to a large extent with the judgement of its appropriateness for decision support.

To be noted is the high discrepancy between the judgements of the individual test users across all criteria. Thus, the κ inter-evaluator agreement measure for ‘comprehensibility’ was only 0.17, for ‘information order’ 0.33, and for ‘content relevance’ 0.13. However, with a simplified grading scale ‘poor’ – ‘acceptable’ – ‘good’, the κ for ‘comprehensibility’ reached 0.49, for ‘information order’ 0.42, and for ‘content relevance’ 0.37.

8 Conclusions

Services that aim at the delivery of publicly relevant information contained in numeric time series still tend to use graphics, tables, pictograms or color scales as the only presentation mode. However, the visual mode is not sufficient if the information is required to have an explanatory dimension, or if it needs to be placed into a larger (user-tailored) context. Advanced text generation technology is then required. Air quality information is such a type of information.

We have presented MARQUIS, a prototypical generator for the generation of multilingual air quality bulletins. Although the generated bulletins still show some deficiencies with respect to both the selection of the communicated content and the linguistic realization, the evaluation experiments demonstrates that the bulletins are well received and judged as being sufficiently well-written.

The contribution of MARQUIS to the field of report generation can be considered twofold. First, it shows the relevance of the addressee (or user) in generation and incorporates a relatively fine-grained user typology. Second, it puts into practice a theoretically motivated and largely domain-independent multilingual generation approach—which is very crucial for the reusability of the system.

The current version of the MARQUIS generator provides a solid basis for our future work towards an off-the-shelf large coverage report generator. In particular, we plan to identify and separate domain-independent and domain-dependent aspects of discourse planning. This will ensure an easier and faster portability of the generator to other applications. Another central strand of our future work will be the acquisition of grammatical and lexical resources using machine learning techniques and the coverage of further languages.

Acknowledgments

MARQUIS has been partially funded by the European Commission under the contract number EDC-11258 in the framework of its R&D programme *eContent*. Many thanks to all partners of the Consortium for the very stimulating and pleasant working atmosphere and, in particular, to those partners who took upon themselves the burden of the evaluation of the reports generated by the MARQUIS service.

References

Bøhler, T., K. Karatzas, G. Peinel, T. Rose & R. San José. 2002. Providing multi-modal access to environmental data-custamisable information services for disseminating urban air quality information in apnee. *Computers, Environment and Urban Systems*. 26.1:39–61.

Bohnet, B., A. Langjahr & L. Wanner. 2000. A Development Environment for MTT-Based Sentence Generators. *Proceedings of the XVI SEPLN Conference*. Vigo, Spain.

Bohnet, B., L. Wanner, R. Ebel, B. Knörzer, M. Tauber & W. Weiß. 2001. Autotext-UIS: Automatische Produktion von Ozonkurzberichten im Umweltinformationssystem Baden-Württemberg. *Proceedings of the Workshop Hypermedia und Umweltschutz*. Ulm.

Bohnet, B. 2006. *Textgenerierung durch Transduktion linguistischer Strukturen*. Berlin: Akademische Verlagsanstalt.

Bronder, J., C. Kliś & J. Długosz. 2007. Air quality information service in Upper Silesia. *Proceedings of the EnviroInfo Conference*. 31–39. Warsaw.

Busemann, S. & H. Horacek. 1997. Generating air-quality reports from environmental data. *Proceedings of the DFKI Workshop on Natural Language Generation*. 15–21. Saarbrücken, Germany.

Caldwell, D. & T. Korelsky. 1994. Bilingual generation of job descriptions from quasiconceptual forms. *Fourt Conference on Applied Natural Language Processing*. 1–6. Stuttgart, Germany.

Cawsey, A.J., R.B. Jones & J. Pearson. 2000. The evaluation of a personalised health information system for patients with cancer. *User Modeling and User-Adapted Interaction*. 10.1:47–72.

Coch, J. 1998. Interactive generation and knowledge administration in MultiMeteo. *Ninth International Workshop on Natural Language Generation*. 300–303. Niagara-on-the-Lake, Ontario, Canada.

Dalianis, H. 1999. Aggregation in natural language generation. *Computational Intelligence*. 15.4:384 – 414.

Ferreira, F., H. Tente, P. Torres, S. Cardoso & J.M. Palma-Oliveira. 2000. Air Quality Monitoring and Management in Lisbon. *Environmental Monitoring and Assessment*. 65:443–450.

Foster, M.E. & M. White. 2004. Techniques of Text Planning with XSLT. *Proceedings of NLPXML'04*. Barcelona.

Goldberg, E., N. Driedger & R. Kittredge. April 1994. Using Natural Language Processing to Produce Weather Forecasts. *IEEE Expert*.

Grell, G.A. 1993. Prognostic evaluation of assumptions used by cumulus parametrizations. *Monthly Weather Revue*. 121:764–787.

Hovy, E.H. 1990. Pragmatics and Natural Language Generation. *Artificial Intelligence*. 43:153–197.

Iordanskaja, L.N., R. Kittredge & A. Polguère. 1991. Lexical Selection and Paraphrase in a Meaning-Text Generation Model. *Natural Language Generation in Artificial Intelligence and Computational Linguistics* ed. by C.L. Paris, W.R. Swartout & W.C. Mann. Kluwer Academic Publishers.

Iordanskaja, L.N., M. Kim, R. Kittredge, B. Lavoie & A. Polguère. 1992. Generation of Extended Bilingual Statistical Reports. *COLING-92*. 1019–1022. Nantes.

Iordanskaja, L. 1992. Communicative Structure and its Use during Text Generation. *International Forum on Information and Documentation*. 17.2:15–27.

Jakobs, H., M. Memmesheimer & A. Ebel. 2005. Daily forecast of air quality over Europe with the EURAD model system. *Geophysical Research Abstracts, Vol. 7, 10739*. European Geosciences Union.

Johansen, P.H., K. Karatzas, J.E. Lindberg, G. Peinel & T. Rose. 2001. Citizen-centered information dissimination with multimodal information channels. *Sustainability in the Information Society. 15th International Symposium on Informatics for Environmental Protection*. Zürich.

Karatzas, K. 2007. State-of-the-art in the dissemination of AQ information to the general public. *Proceedings of the EnviroInfo Conference, Volume 2*. 41–47. Warsaw.

Kittredge, R.I. & A. Polguère. 2000. The Generation of Reports from Databases. *Handbook of Natural Language Processing* ed. by R. Dale, H. Moisl & H. Somers. 261–304. New York: Taylor and Francis.

Kittredge, R., T. Korelsky & O. Rambow. 1991. On the need for domain communication knowledge. *Computational Intelligence*. 7.4:305 – 314.

Kukich, K. 1983. Knowledge-Based Report Generation: A Technique for Automatically Generating Natural Language Reports from Databases. *Proceedings of the Sixth International ACM SIGIR Conference*. Washington, DC.

Kukkonen, J., J. Härkönen, J. Walden, A. Karppinen & K. Lusa. 2001. Evaluation of the dispersion model CAR-FMI against data from a measurement campaign near a major road. *Atmospheric Environment*. 35.5:949–960.

Lambrecht, K. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press.

Lareau, F. & L. Wanner. 2007. Towards a generic multilingual dependency grammar for text generation. *Proceedings of the GEAF07 Workshop* ed. by T. King & E.M. Bender. 203–223. Stanford, CA: CSLI.

Lohmeyer, A., I. Düring, M. Giereth, M. Klein, D. Nicklaß, H. Scheu-Hachtel, C. Sörgel & L. Wanner. 2007. Kurzfrist-Feinstaub-Immissionsprognose mit den Systemen MARQUIS und ProFet. *Gefahrstoffe–Reinhaltung der Luft*. 7/8:319–326.

Mann, W.C. & S.A. Thompson. 1987. Rhetorical Structure Theory: A theory of text organization. *The Structure of Discourse* ed. by L. Polanyi. Norwood, New Jersey: Ablex Publishing Corporation.

McKeown, K. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge, England: Cambridge University Press.

Mel'čuk, I.A. 1988. *Dependency Syntax: Theory and Practice*. Albany: SUNY Press.

Mel'čuk, I.A. 1996. Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. *Lexical Functions in Lexicography and Natural Language Processing* ed. by L. Wanner. 37–102. Amsterdam/Philadelphia: Benjamins Academic Publishers.

Mel'čuk, Igor A. 2001. *Communicative Organization in Natural Language (The Semantic-Communicative Structure of Sentences)*. Amsterdam: Benjamins Academic Publishers.

Molina, T., A. Panighi & A. Flores. 2005. MARQUIS (EDC-11258), Deliverable 3.1: Specification of the User Profiles. Technical report: TVC Netmedia.

Neto, J., P. Torres, F. Ferreira & F. Boavida. 2005. Lisbon air quality forecast using statistical methods. *Proceedings of the AQM 2005 conference in Istanbul*. 591–597.

Oglesby, L., N. Künzli, M. Rööli, Braun-Fahrländer, P. Mathys, W. Stern, M. Jantunnen & A. Kousa. 2000. Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution. *Journal of the Air Waste Management Association*. 50:1251–61.

Paris, C. 1993. *User Modelling in Text Generation*. London: Frances Pinter Publishers.

Peinel, G., T. Rose & R. San José. 2000. Customized Information Services for Environmental Awareness in Urban Areas. *Proceedings of the Seventh World Congress on Intelligent Transport Systems*. Turin.

Polguère, Alain. 1990. *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cardre de génération de texte*. PhD thesis: Département de linguistique, Université de Montréal.

Polguère, A. 1998. Pour un modèle stratifié de la lexicalisation en génération de texte. *t.a.l.* 39.2:57–76.

Portet, F., E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer & C. Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*. 173.7-8:789–916.

Rösner, D. & M. Stede. 1992. Customizing RST for the Automatic Production of Technical Manuals. *Aspects of Automated Natural Language Generation* ed. by R. Dale, E. Hovy, D. Rösner & O. Stock. Berlin: Springer Verlag.

Reiter, E. 1994. Has a Consensus in NL Generation Architecture Appeared, and is it Psycholinguistically Plausible? *Proceedings of the 7th International Workshop on Natural Language Generation*. 163–170. Kennebunkport.

Rösner, D. 1986. *Ein System zur Generierung von deutschen Texten aus semantischen Repräsentationen*. PhD thesis. Stuttgart, Germany: Institut für Informatik, Stuttgart University.

Sowa, J. 2000. *Knowledge Representation*. Pacific Grove, CA: Brooks Cole.

Sripada, S., E. Reiter & I. Davy. 2003. SumTime-Mousam: Configurable marine weather forecast generator. *Expert Update*. 6.3:4–10.

Vallduví, E. 1995. Information packaging: A survey. Report: Center for Cognitive Science and HCRC, University of Edinburgh.

Van Loon, M., M.G.M Roemer & P.J.H. Builtjes. 2004. Model inter-comparison in the framework of the review of the unified EMEP model. Technical Report TNO Report R2004/282: TNO.

Walker, M., S. Whittaker, A. Stent, P. Maloor, J. Moore, M. Johnston & G. Vasireddy. 2004. Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science*. 28.5:811–840.

Wanner, L., D. Nicklaß, B. Bohnet, N. Bouayad-Agha, J. Bronder, F. Ferreira, R. Friedrich, A. Karppinnen, F. Lareau, A. Lohmeyer, A. Panighi, S. Parisio, H. Scheu-Hachtel & J. Serpa. 2007. From Measurement Data to Environmental Information: MARQUIS–A Multimodal

Air Quality Information Service for the General Public. *Proceedings of the 6th International Symposium on Environmental Software Systems* ed. by A. Swayne & J. Hrebicek. Prague.

Wanner, L. 1996. Introduction. *Lexical Functions in Lexicography and Natural Language Processing* ed. by L. Wanner. 1–36. Amsterdam: Benjamins Academic Publishers.

Wanner, L. 1996. Lexical Choice in Text Generation and Machine Translation. *Machine Translation.* 11.1–3.

Wilcock, G. 2001. Pipelines, Templates and Transformations: XML for Natural Language Generation. *Proceedings of NLP XML-2001.*

Yu, J., E. Reiter, J. Hunter & C. Mellish. 2007. Choosing the Content of Textual Summaries of Large Time-Series Data Sets. *Natural Language Engineering.* 13.1:25–49.

Zukerman, I. & D. Litman. 2001. Natural language processing and user modeling: Synergies and limitations. *User Modeling and User-Adapted Interaction.* 11:129–158.

Zukerman, I. & R. McConachy. 1994. Discourse Planning as an Optimization Process. *Proceedings of the 7th International Workshop on Natural Language Generation, 3-6 June 1990.* 37–44. Kennebunkport, MA.