Issues in Generating Text from Interlingua Representations

Stephan Busemann

DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
`busemann@dfki.de`

Abstract

Multi-lingual generation starts from non-linguistic content representations for generating texts in different languages that are equivalent in meaning. In contrast, cross-lingual generation is based on a language-neutral content representation which is the result of a linguistic analysis process. Non-linguistic representations do not reflect the structure of the text. Quite differently, language-neutral representations express functor-argument relationships and other semantic properties found by the underlying analysis process. These differences imply diverse generation tasks. In this contribution, we relate multi-lingual to cross-lingual generation and discuss emergent problems for the definition of an interlingua.

1. Introduction

In this contribution, we relate multi-lingual to cross-lingual generation and discuss emerging problems for the definition of an interlingua. Multi-lingual generation starts from non-linguistic content representations for generating texts in different languages that are equivalent in meaning. The generation of weather forecasts or environmental reports are typical examples. In contrast, cross-lingual generation is based on a language-neutral content representation which is the result of a linguistic analysis process. Generation for machine translation is a most prominent example.

Non-linguistic representations do not specify linguistic semantics nor do they reflect the structure of the text to be generated. In contrast, language-neutral representations express functor-argument relationships and other semantic properties found by the underlying analysis process. These differences imply diverse generation tasks.

However, there are also commonalities. In both cases, generation is the mapping of some semantic representation onto linguistic strings. We may assume a single generation process that uses different separately defined language-specific knowledge sources. In both cases, we may view the underlying representation as an interlingua, since it attempts to cross the language barrier by providing content descriptions independently of the target language.

An instance of each type of tasks has been implemented using the generation system TG/2 (Busemann, 1996), quickly overviewed in Section 2.. The usage of the same framework allows us to relate the tasks to each other (Section 3.) and to gain insights relevant to a coherent definition of interlinguas, generation tasks, and generation knowledge (Section 4.).

2. TG/2 in a Nutshell

TG/2 is a flexible production system that provides a generic interpreter to a set of user-defined condition-action rules representing the generation grammar. The generic task is to map an input structure onto a chain of terminal elements as prescribed by the rule set. The rules have a context-free categorial backbone used for standard top-down derivation, which is guided by the input representation. The rules specify conditions on input ("tests") determining their applicability and allow navigation within the input structure ("access functions").

The right-hand side of a rule can consist of any mixture of terminal elements (canned text) or other categories associated with an access function. The presence of canned text is useful if the input does not express explicitly everything that should be generated. With very detailed input, the terminal elements of the grammar will usually be words.

Given a category C and some (piece of) input structure I, production rules are applied through the standard three-step processing cycle:

1. Identify the applicable rules;

2. Select a rule on the basis of some (freely programmable) conflict resolution mechanism; and

3. Apply that rule.

A rule is applicable if its left-hand side category is C and its tests hold on I. A rule is applied by processing its right-hand side elements from left to right. Canned text is output right away, and non-terminal elements induce a new cycle with the new category and the return value of the access function. Processing terminates when all right-hand side elements have been realized successfully. In the case of a failure, processing backtracks to step 2. If no more rules are applicable, a global failure occurs. For details see (Busemann, 1996).

3. Relating Two Distinct Generation Tasks

TG/2 has been used in a variety of NLG tasks. We look at multi-lingual report generation and cross-lingual summarization. We then locate the tasks on a scale ranging from shallow to in-depth generation, and discuss advantages and drawbacks of these locations.

3.1. Task 1: Generating air quality reports from measurement data

Reports about air quality in a German-French border region (Busemann and Horacek, 1998) are currently

```
[(COOP THRESHOLD-PASSING)
 (LANGUAGE ENGLISH)
 (TIME [(PRED SEASON)
        (NAME [(SEASON WINTER)
               (YEAR 2001)])])
 (POLLUTANT SO2)
 (SITE "Saarbruecken-City")
 (SOURCE [(THRESHOLD-TYPE MIK-WERT)])
 (EXCEEDS [(STATUS YES) (TIMES 1)])]
```

Figure 1: A Non-Linguistic Input Expression for Report Generation: "In Winter 2001 at the measuring station at Saarbrücken-City, the MIK value for sulfur dioxide was exceeded once."

```
(defproduction site "S01"
  (:PRECOND
    (:CAT SITE-E
     :TEST ((always-true)))
  :ACTIONS
    (:TEMPLATE
      "at the measuring station at "
      (:RULE SITE-NAME-E (self)))))
```

Figure 2: Making Implicit Meaning Explicit: A TG/2 grammar rule. The rule is "unconditioned" and uses the current piece of input structure to access the site name.

produced in six languages (a web demo is available at `http://www.dfki.de/service/nlg-demo`). The reports are based on real measurement data taken from a database and on the user's parameters determining the type of the report (time series, average or maximum value description, threshold passing description). A report consists of up to six statements most of which are verbalized by TG/2. The initial text organization stage retrieves the relevant data, decides about the content of the statements and defines their order. For each statement to be verbalized by TG/2 it produces a domain-oriented non-linguistic intermediate feature structure serving as input to TG/2 (cf. Figure 1 for an example). Input expressions for TG/2 may specify e.g. the pollutant, the actual measurements, and their date and location. Moreover, further information is specified according to the user's choice of parameters. It should be noted that some input is just carried forward from the original system input (in Figure 1, this is LANGUAGE, TIME, POLLUTANT, SITE, THRESHOLD-TYPE), whereas other information originates from the DB query and text organization stage (COOP and EXCEEDS in Figure 1).

The text organization stage is entirely content-oriented, and the intermediate feature structures do not exhibit linguistic properties. The 'language' feature causes the selection of the rule set for the language requested. The determination of linguistic structure for each input expression is achieved by the TG/2 grammar rules. Since implicit information is associated with some parts of input expressions, canned text is used to make it explicit at the surface. An example in Figure 1 is the added notion of "at the measuring station at" in the case of (SITE "Saarbrücken-City"), which is verbalized through the rule in Figure 2.

The grammars comprise about 100-120 rules for each language and are specifically designed for this application. The development of a grammar for another language takes between one and three weeks depending on skills.

3.2. Task 2: Generating medical scientific text for summaries

This generation task occurred in the context of the cross-lingual text summarization system MUSI (Lenci et al., 2002). MUSI involves a combination of analysis and generation similar to machine translation. An interlingua approach was chosen to represent selected English and Italian medical scientific sentences in a language-neutral way. The sentences can be complex and quite long (50 words are no exception). Interlingua expressions were fed to sentence generation components producing the elements of a French or German summary.

The generation of German sentences (Busemann, 2002) starts from so-called IRep4 interlingua expressions. A sample IRep4 expression is shown in Figure 3. IRep4 expressions are hierarchical predicate-argument structures complemented by a rich variety of features and modifiers. The basic elements are atomic and predicative concepts, forming an ontology shared across the MUSI system. In particular, predicative frames are based on the SIMPLE formal specifications (Lenci et al., 2000). IRep4 expressions are composed of PROP and ITEM elements used to represent propositions and terms, respectively. Although IRep4 is in principle a semantic representation language, its expressions also keep track of some syntactic properties of the source language elements. For instance, number and determiner information is specified for NPs as well as categorial information for propositions (CAT). This information can be very useful in guiding text generators.

IRep4 is suitable for representing the semantics of very complex sentences, but at the same time, it leaves room for various degrees of specification. In fact, co-reference resolution, attachment ambiguities and the incorrect identification of arguments and modifiers are common sentence analysis problems that may lead to incomplete output. To cope with these problems, IRep4 has been designed to integrate possibly underspecified or fragmentary representations. This feature greatly enhances the robustness of the system and can guarantee a better interface with the text analysis component.

A direct interpretation of IRep4 by TG/2 would require choosing the lexemes and the syntactic realizations. This could have been achieved within the TG/2 grammar through complicated tests. These choices partly depend on each other, which would have caused massive backtracking. Moreover, testing the presence of a concept in IRep4 would have been triggered by rules expanding the syntactic category of the lexemes (part of speech), e.g. the rule Noun → "acetylcholin" would have been associated with a test whether the current concept was C_acetylcholine. As there would have been hundreds of these, concerns of processing efficiency were in order. Finally, a pre-existing grammar should be reused that was not previously adapted

```
PROP{ Value = P_ARG1_cause_ARG2;
      Time_Rep = [PRESENT, PRES_USUAL];
      Cat = V_SEN;
      Arg1 = PROP{ Value = P_antagonism_with_ARG1;
                     Cat = NP; Det = INDEF;
                     Arg1 = ITEM{ Value = C_acetylcholine;
                                   Mod1 = [LOC, ITEM{
                                             Value = C_level;
                                             Det = DEF;
                                             Mod1 = [RESTR, ITEM{
                                                       Value = C_sight;
                                                       Number = PLUR; Det = DEF;
                                                       Mod1 = [RESTR, C_muscarinic];
                                                       Mod2 = [RESTR, ITEM{
                                                                 Value = C_substance;
                                                                 Number = PLUR;
                                                                 Det = DEMONST1;}]; }]; }]; };
                     Mod1 = [RESTR, C_competitive]; };
      Arg2 = ITEM{ Value = C_effect;
                     Det = DEF; Number = PLUR; }; }
```

Figure 3: IRep4 Expression for "Die Wirkungen werden durch einen kompetitiven Antagonismus zu Acetylcholin auf dem Niveau der muskarinischen Bindungsstellen dieser Substanzen verursacht." [The effects are caused by a competitive antagonism with acetylcholine on the level of the muscarinic sights of these substances.].

to IRep4.

For these reasons it appeared more convenient to introduce an initial sentence planning stage. The resulting representation – see Figure 4 for an example corresponding to Figure 3 – forms the input to TG/2. It can be viewed as a syntactically enriched, language-specific paraphrase of the underlying IRep4 expression. It represents explicitly the linguistic structure of the sentence. The TG/2 grammar is responsible for word order and inflection. Very much like in a classical sentence realization system, no canned text parts are used. If a phrase like "at the measuring station at" had to be generated here, an underlying interlingual semantic expression would be mandatory.

A pre-existing TG/2 grammar for German syntax was reused and adapted to the needs of MUSI (Busemann, 2002; Lenci et al., 2002). Its final version comprises over 950 rules.

3.3. Shallow and in-depth generation

The notion of shallow generation, as opposed to in-depth generation, has been coined by (Busemann and Horacek, 1998) to describe a distinction corresponding to that of shallow and deep analysis. In language understanding deep analysis attempts to "understand" every part of the input, while shallow analysis tries to identify only parts of interest for a particular application, omitting others. In-depth generation is inherently knowledge-based and theoretically motivated, whereas shallow generation quite opportunistically models only the parts of interest for the application in hand. Often such models will turn out to be extremely shallow and simple, but in other cases much more detail is required. Thus, techniques such as those developed within TG/2 for varying modeling granularity according to the requirements posed by the application are a prerequisite for reusing NLG systems.

Obviously a shallow NLG system is, in general, based on representations that carry implicit meaning. We call this shallow input. Additional text has to be "invented" by the generator (in TG/2, this is usually achieved using canned text in the grammar).[1] This leads to domain-dependent, shallow grammars that cannot be reused easily for another task. The in-depth models assume a very fine-grained grammar describing all the linguistic distinctions covered by the interlingua. Such a grammar corresponds closely to familiar generic linguistic resources.

The report generation task described was solved by a typical shallow approach, whereas the MUSI generation task required an in-depth model.

The tension between shallow and in-depth generation has been discussed further in the literature. According to Reiter and Mellish, shallow techniques (which they call "intermediate") are appropriate as long as corresponding in-depth approaches are poorly understood, less efficient, or more costly to develop (Reiter and Mellish, 1993). Bateman and Henschel describe ways of compiling specialized grammars out of general resources (Bateman and Henschel, 1999). A platform for generating, storing and reusing representations is described in (Calder et al., 1999), showing that such reuse can be seen as a shallow methodology to text generation. A major conclusion seems that there is no dichotomy between both approaches, but that shallow systems can indeed be based on theoretically sound in-depth models.

In practice though, NLG tasks turn out to be highly diverse, and no NLG system could be reused for a new application off the shelf. The necessary effort for adaptation and extension of large existing in-depth resources such as KPML (Bateman, 1997) or FUF/Surge (Elhadad and Robin, 1996) is often considered high. In fact, the de-

---

[1]Of course, these texts are defined by the application, viz. the customer, as all other output.

```
[(SENTENCE DECL)
 (VC [(VOICE PASSIV)
      (MOOD IND)
      (TENSE PRAESENS)
      (SBP S2)
      (STEM "verursach")])
 (DEEP-SUBJ [(TOP Y)
             (TY GENERIC-NP)
             (NUMBER SG)
             (DET INDEF)
             (NR V2)
             (GENDER MAS)
             (STEM "antagonismus")
             (PP-ATR [(LOCATIVE ...)
                      (GENDER NTR)
                      (STEM "Acetylcholin")
                      (DET WITHOUT)
                      (NUMBER SG)
                      (TY GENERIC-NP)
                      (PREP MIT)])
             (ADJ [(STEM "kompetitiv")
                   (POS ADJECTIVE)
                   (DEG POS)])])
 (DEEP-AKK-OBJ [(TY GENERIC-NP)
                (NUMBER PLUR)
                (DET DEF)
                (STEM "wirkung")
                (GENDER FEM)])]
```

Figure 4: TG/2 Input Expression Partly Corresponding to Figure 3. The material for "on the level of the muscarinic sights of these substances" would appear under `DEEP-SUBJ.PP-ATR.LOCATIVE`, but has been omitted for reasons of space. The representation contains content word stems and names for syntactic structures (`SBP, NR` features). Determiners and prepositions are also provided.

velopment from scratch of a shallow grammar for a small NLG application on the basis of a simple framework like TG/2 can be more cost-effective.

Shallow and in-depth generation tasks can be related with help of TG/2. As the amount of domain-specific canned text in the TG/2 grammars correlates to the shallowness of the input, the generation tasks described can be located on a scale that ranges from shallow to in-depth domain and input models. There are trivial systems at one end that just produce canned text according to triggers (e.g. system error reports). A bit further on the scale we find template-style systems, like the air quality report generator, which use canned text to make knowledge implicit in the input explicit. In-depth realizers with sophisticated grammars that do not use domain-specific canned text at all are located at the other end of the scale, such as the MUSI generator.

Why are shallow and in-depth interlinguas both viable? One obvious reason lies in the origin of the interlingua representations. Shallow representations usually originate from non-linguistic processing, such as accessing a database or interpreting some user interaction, whereas in-depth representations generally have a linguistic origin, e.g. from an NL parsing component.

More interestingly, the type of domain and application determines the depth of modeling. Air quality reports form a small and closed domain. Implicit knowledge is easy to make explicit. A shallow model, being inherently simple, is perfectly adequate. A complex functor-argument representation would mean a dramatic overshot for this type of application. The same holds for many generation applications, such as reporting about stock exchange (Kukich, 1983) or weather forecasts (Boubeau et al., 1990). Medical scientific texts, on the other hand, form a very large domain, requiring broad-coverage linguistic knowledge. A shallow model would not even be able to capture the most frequent semantic relations. General means of expressing semantic relationships are mandatory.

What are the advantages and drawbacks of either approach? Shallow interlinguas allow for a straightforward multi-lingual generation. All linguistic processing can be concentrated in the module consuming the interlingua expression, e.g. TG/2. A drawback consists in domain-dependent grammars, which are hardly reusable for other applications. Still it is worthwhile, as the effort to create a grammar for another language is low.

With in-depth language-neutral representations, the issue of reusing existing linguistically motivated grammars arises, simply because of the tremendous effort for developing them from scratch. Technically an existing grammar may be reused if a well-defined interface is available. In TG/2, the interface to the input representations consists of the tests and access functions called from within the grammar rules. Depending on the different organization of information within input languages, this interface must be modified. If the same types of information required by the grammar can be produced by the new input language, the way is paved for a successful reuse. If the new input language offers different types of information, the adaptation problem described above arises.

4. On the Definition of Interlinguas

We now address issues on the semantics and pragmatics of interlinguas from a generation perspective by discussing three types of problems generators may encounter with in-depth interlinguas, using experiences with IRep4 as our source of examples.[2]

4.1. Extrinsic problems

In MUSI, a variety of problems with interlinguas known from machine translation were experienced, showing that this interlingua, as so many others, is not language-neutral in a strict sense. The problems were related to the fact that languages encode information differently and the interlingua cannot sufficiently abstract away from this. More precisely, although IRep4 does not contain elements specific to any of the four languages involved, the analysis results reflected some grouping and nesting of phrases and clauses of the source language.

---

[2] By critically reviewing IRep4, we necessarily omit mentioning many excellent features that made it very useful for the challenging task of representing scientific text.

For instance, Italian (and English) uses post-nominal adjectival clauses that correspond to a post-nominal relative clause or pre-nominal adjectival modifiers in German (cf. Figure 5a). German does not have the possibility to linearize or nest several adjectival or participial clauses after the head noun. Moreover, large phrases in pre-nominal position are difficult to understand since the head noun is uttered only afterwards.

In IRep4, these clauses are typically represented as restrictive modifiers (RESTR), accompanied, in the case of a predicative concept, by the source-language specification CAT = ADJP. The generator follows the heuristic strategy of assigning small adjectival phrases to the pre-nominal adjective position and large ones to the post-nominal relative clause position. In the latter case, the CAT specification will be ignored, as a full sentence with a copula must be generated. A further requirement consists of the need for one argument of the adjective to be realizable as the relative pronoun.

The result is not satisfactory, as it can lead to recursive center-embedding causing bad readability (cf. Figure 5b). The sentence in Figure 5c is stylistically much better; it has fewer closing brackets in a sequence, which means less deep embedding and improved readability. Linguistically, it shows two extrapositions, i.e. the innermost relative clause (not bracketed further) occupies the post-field[3] of the embedding one, which in turn occupies the post-field of the main clause. The stylistically preferred solution would be to realize the innermost clause as a prenominal AP, while extraposing the larger clause as a relative clause, as in Figure 5d.

Another striking example of language differences experienced with IRep4 is the use of determiners. English text does not use always definite articles when they are mandatory in German. For instance, "features of malnutrition" should be translated into "Merkmale der Mangelernährung" (definite article included), whereas "features of chronic malnutrition" corresponds to "Merkmale chronischer Mangelernährung" (no article).

IRep4 does, of course, not represent definite articles when there are no such determiners in the source-language text. The generator uses as a general rule that "naked" generalized possessives – i.e. the head of a RESTRictive modifier that corresponds to a noun and does not have a determiner or a modifier – are automatically accompanied by a definite article, covering the above examples.

English "Treatment consisted in..." should translate to "Die Behandlung bestand aus...", using a definite article. In these cases, a decision within the generator on whether or not to use a definite article would rely on lexical semantic information about both the source and target language lexemes.

The obvious solution to the extrinsic problems is to complement the level of interlingua with a set of transfer rules specific for every pair of source and target language. This complicates the situation, but would, in MUSI, have

led to considerable stylistic improvements of the generated sentences.

For shallow models, this problem simply does not exist.

4.2. Intrinsic problems

IRep4 also has a few intrinsic properties that affected generation. Most prominently, it does not represent scope and thematic, or constituent, order information. The scope of negation would be important for the proper placement of the negation particle. Moreover, the scope of modifiers is not represented. With the current, inherently flat representation, i.e. multiple modifiers at the same level of embedding, generation cannot decide between e.g. "the following clinical case" and "the clinical following case". Modifiers should be nested to express this information.

Deciding about word order in generation is relevant to represent the argumentative structure in complex sentences and ensure coherence. The order of constituents in the source language text is not marked in IRep4, which may cause a deviating target-language order in German. This can lead to a lack of textual coherence, if e.g. a modifier that starts the sentence appears at the end. Consider "upon objective investigation, the woman's face was red and congested", which was translated into "das Gesicht der Frau war rot und geschwollen bei objektiver Untersuchung", generating the introductory PP at the end. A possible subsequent anaphoric reference would be less felicitous than in the original text. In the absence of a super-ordinated text planning stage, interlingua expressions should specify thematic order, or constituent order, in the source language text.

German generation assumes a standard word order for active voice, unless other information is given. The standard word order does not take into consideration the complexity, or the "weight", of a constituent. A heavy-weight subject preceding a short object in a transitive sentence is often considered bad style. Based on heuristics about a constituent's "weight", passive voice could have been chosen within the generator, causing the short constituent to precede the complex one, which generally leads to more fluent text (cf. the example in Figure 3). An interlingua should include hooks to provide this information. IRep4 might indirectly allow a good estimate by counting concepts, arguments and modifiers; further investigation is needed to identify a reliable formula.

For shallow interlinguas, intrinsic problems of this kind do not exist, as they are entirely dealt with in the grammar.

4.3. Pragmatic problems

In this section, we sketch some issues that can take a lot of effort to create a shared understanding among the researchers looking at interlingua expressions from different perspectives.

A grammatically correct input sentence is a legitimate input to a parser. Few systems can deal with incorrect sentences in an error-tolerant way. For generation, in-depth interlingua expressions should be correct in a similar sense. A formal specification of the interlingua is required to define its syntax and, very importantly, its semantics. Generation requirements should be formally specified as well and

---

[3]The post-field follows the infinite verb complex in a German declarative sentence. This position can be occupied by one constituent.

a) [[In the clinical case described,] [the symptoms] [were] [caused] [by ingestion [of anticolinergic substances [probably contained [in the leaves [of plants [consumed a few hours before]]]]]]].

b) [[In dem beschriebenen klinischen Fall] [wurden] [die Symptome] [durch [Verzehr [von anticholinergen Substanzen, [[die] [die Blätter [der Pflanze], [die vor ein paar Stunden genossen wurden,] möglicherweise enthielten,]]]]]] [verursacht]].

In the described clinical case were the symptoms by ingestion of anticolinergic substances, that-were in-the leaves of-the plants, that-were a few hours before consumed, possibly contained.

c) [[In dem beschriebenen klinischen Fall] [wurden] [die Symptome] [durch Verzehr [von anticholinergen Substanzen]] [verursacht], [[die] [die Blätter [der Pflanze]] möglicherweise enthielten, [die vor ein paar Stunden genossen wurden]]].

d) [[In dem beschriebenen klinischen Fall] [wurden] [die Symptome] [durch Verzehr [von anticholinergen Substanzen]] [verursacht], [[die] [die [vor ein paar Stunden genossenen] Blätter [der Pflanze]] möglicherweise enthielten]].

Figure 5: Stylistic Variations in Translation. Brackets indicate some syntactic structure. a) English original sentence; b) Corresponding sentence in German with APs realized as relative clauses, with inter-linear translation; c) Extraposition of the relative clauses beyond the respective verbs; d) Realization of the innermost clause as a prenominal AP.

should be part of the "pragmatics" of the interlingua. For instance,

- the omission of information about tense, aspect, determination and number may mean that a default applies;

- a personal pronoun must either refer to an antecedent, or be accompanied by information about gender, person and number;

- an expression realized as a relative clause must contain exactly one constituent with a plain coreference specification; this constituent will become the relative pronoun;

- etc.

During the development of IRep4, this effort was not spent due to shortage of resources.[4] While from an analysis viewpoint, some decent output looks more or less satisfactory, it is the details that make generation feasible or cause its failure. Most importantly, the interpretation of interlingua expressions in NLG should be functional. Different surface representations corresponding to the same interlingua expression should be considered as equivalent in meaning. If this fundamental principle is not maintained, translation is not guaranteed to be meaning-preserving.

An interlingua can support this principle by making meaning representation explicit. IRep4 unfortunately has a fairly abstract representation for PP adjuncts and modifiers. The scheme is "Mod = [<name>, <Irep4-expression>]", where <name> is taken from a finite set of strings that more or less denote the semantics of the modifier. These names can be interpreted unambiguously by generation, but analysis may encounter difficulties in relating prepositions and head nouns to them, if only little lexical semantic knowledge is available. In Figure 3, the same name RESTR is realized differently, depending

on the part of speech used for the embedded concept. If it is a noun, the semantics is that of a generalized possessive, which is realized in post-nominal position in German. If it is an adjective, a prenominal adjectival modifier is usually generated. Other uses of RESTR were mentioned above. If two or more meanings are connected to one name, it may appear psychologically difficult to refrain from using this name as a waste-basket.

Pragmatic problems exist for shallow models as well, as shallow input expressions are partly produced by external systems. In the air quality report generator, measuring values are received as input from a database. Time series are occasionally shortened by aggregating information ("from 9.00 to 11.00: 6,7 μg/m^3"). During the development, we have not been aware of the systematic omission of certain half hour values in the database, which occasionally leads to awkward results: "at 9.00: 6,7 μg/m^3; at 9.30: 0 μg/m^3; at 10.00: 6,7 μg/m^3; at 10.30: 0 μg/m^3; at 11.00: 6,7 μg/m^3". We easily could have implemented another aggregation rule that leads to output like "from 9.00 to 11.00: 6,7 μg/m^3, with every half hour value at 0".

5. Conclusion

In this contribution, we have related multi-lingual to cross-lingual generation and discussed emerging problems for the definition of an interlingua. This discussion was based on experience gained from implementing NLG components for a multi-lingual report generator and a cross-lingual summarization system within the same framework, TG/2. Shallow interlinguas originate from non-linguistic processing. They usually carry implicit meaning that must be made explicit in the generation process. For relatively small-coverage, closed domains, such as air quality reports, weather reports, or stock market reports, it is adequate to write specialized grammars using domain-specific canned text for this purpose. In-depth interlinguas usually originate from linguistic analysis, as in machine translation. The nature of the interlingua is closely tied to the sophistication of

---

[4]It is debatable though whether the resulting difficulties have been resolved with less effort.

the generation task in hand.

While well-modularized generation systems can be easily adapted to shallow interlinguas, an in-depth interlingua is much more complex to work with, as so many distinctions need to be addressed. In this paper we have identified some NLG requirements on in-depth interlinguas. From the experience with the MUSI application, we have learned that it is worthwhile to formally specify NLG requirements on the interlingua at the outset.

For a new application involving multi-lingual or cross-lingual generation, the interlingua should be chosen, adapted or designed according to the kind of linguistic processing involved and in view of the depth of modeling envisaged. On the shallow/in-depth scale, it should be as shallow as possible.

6. References

John Bateman and Renate Henschel. 1999. From full generation to 'near-templates' without loosing generality. In (Becker and Busemann, 1999), pages 13–18. Also available at `http://www.dfki.de/service/NLG/KI99.html`.

John Bateman. 1997. KPML delvelopment environment: multilingual linguistic resource development and sentence generation. Report, German National Center for Information Technology (GMD), Institute for integrated publication and information systems (IPSI), Darmstadt, Germany, January. Release 1.1.

Tilman Becker and Stephan Busemann, editors. 1999. *May I Speak Freely? Between Templates and Free Choice in Natural Language Generation. Workshop at the 23rd German Annual Conference for Artificial Intelligence (KI '99). Proceedings*, Document D-99-01. Also available at `http://www.dfki.de/service/NLG/KI99.html`.

L. Boubeau, D. Carcagno, E. Goldberg, Richard Kittredge, and A. Polguére. 1990. Bilingual generation of weather forecasts in an operations environment. In *Proceedings of the 13 th International Conference on Computational Linguistics (COLING-90), Volume 1*, pages 90–92, Helsinki.

Stephan Busemann and Helmut Horacek. 1998. A flexible shallow approach to text generation. In Eduard Hovy, editor, *Nineth International Natural Language Generation Workshop. Proceedings*, pages 238–247, Niagara-on-the-Lake, Canada. Also available at `http://xxx.lanl.gov/abs/cs.CL/9812018`.

Stephan Busemann. 1996. Best-first surface realization. In Donia Scott, editor, *Eighth International Natural Language Generation Workshop. Proceedings*, pages 101–110, Herstmonceux, Univ. of Brighton, England. Also available at the Computation and Language Archive at `http://xxx.lanl.gov/abs/cmp-lg/9605010`.

Stephan Busemann. 2002. Language generation for cross-lingual document summarisation. In Huanye Sheng, editor, *International Workshop on Innovative Language Technology and Chinese Information Processing (ILT&CIP-2001), April 6-7, 2001, Shanghai, China*, Beijing, China, May. Science Press, Chinese Academy of Sciences.

Jo Calder, Roger Evans, Chris Mellish, and Mike Reape. 1999. "free choice" and templates: how to geth both at the same time. In (Becker and Busemann, 1999), pages 19–24. Also available at `http://www.dfki.de/service/NLG/KI99.html`.

Michael Elhadad and Jacques Robin. 1996. An overview of SURGE: a reusable comprehensive syntactic realization component. In Donia Scott, editor, *Eighth International Natural Language Generation Workshop. Demonstrations and Posters*, pages 1–4, Herstmonceux, Univ. of Brighton, England.

Karen Kukich. 1983. Design and implementation of a knowledge-based report generator. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 145–150, Cambridge, MA.

Alessandro Lenci, Nuria Bel, F. Busa, Nicoletta Calzolari, E. Gola, M. Monachini, Alexandre. Ogonowsky, I. Peters, W. Peters, N. Ruimy, M. Villegas, and Antonio Zampolli. 2000. SIMPLE: a general framework for the development of multilingual lexicons. *International Journal of Lexicography*, 13(4):249–263.

Alessandro Lenci, Ana Água, Roberto Bartolini, Stephan Busemann, Nicoletta Calzolari, Emmanuel Cartier, Karine Chevreau, and José Coch. 2002. Multilingual summarization by integrating linguistic resources in the MLIS-MUSI project. In *Procs. Third International Conference on Language Resources and Evaluation (LREC)*, Las Palmas, Canary Islands, Spain, May.

Ehud Reiter and Chris Mellish. 1993. Optimizing the costs and benefits of natural language generation. In *Proc. 13th International Joint Conference on Artificial Intelligence*, pages 1164–1169, Chambery, France.