Meta

Journal des traducteurs Translators' Journal

AT3M

The Realization of Individual Instances in a Multilingual Generation System

Guy Everaert and Thierry van Steenberghe

Volume 39, Number 1, mars 1994

La traduction et l'interprétation dans la Belgique multilingue

URI: https://id.erudit.org/iderudit/002712ar DOI: https://doi.org/10.7202/002712ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print) 1492-1421 (digital)

Explore this journal

Cite this article

Everaert, G. & van Steenberghe, T. (1994). The Realization of Individual Instances in a Multilingual Generation System. Meta, 39(1), 194-205. https://doi.org/10.7202/002712ar

Article abstract

This paper presents multilingual sentence generation in the GENESE project, part of the efforts to develop pivot-language based machine translation. The handling of individual instances will be developed and the problem of mass expressions will be highlighted. This will illustrate the use of available linguistic theories for the generation of individual instances in GENESE.

Tous droits réservés © Les Presses de l'Université de Montréal, 1994

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/



THE REALIZATION OF INDIVIDUAL INSTANCES IN A MULTILINGUAL GENERATION SYSTEM

GUY EVERAERT AND THIERRY VAN STEENBERGHE Université Catholique de Louvain, Institute of Linguistics, Leuven, Belgium

Résumé

Cet article présente le projet GENESE (GENERATION SEMANTIC). GENESE est un système de génération de phrases multilingue et fait parti des expériences de recherche sur le développement du langage-pivot. Dans une première partie on décrit les problèmes reliés à la manipulation des phrases individuelles et en seconde partie on s'attarde aux problèmes des expressions de masse.

This paper presents multilingual sentence generation in the GENESE project¹, part of the efforts to develop pivot-language based machine translation.

The handling of individual instances will be developed and the problem of mass expressions will be highlighted. This will illustrate the use of available linguistic theories for the generation of individual instances in GENESE.

GENERATION IN GENESE

The aim of the GENESE project consists in generating sentences in different languages, specifically French and Dutch, from language independent semantic representations: GENESE stands for *GENEration SEMantiques*.

The input to the generation process is a representation in the form of semantic networks made up of concept nodes linked by relational arcs. Such networks can be rewritten in a linearly ordered manner as follows:

In this representation, ASSERT indicates that the speaker performs an assertional illocutionary act that bears on an event e.

The latter is an occurrence of the concept type *look*: the types are defined in a concept lexicon, that is a heterarchy or lattice network based on the object-attribute-value paradigm, with inheritance rules.

This event is specified by two arguments, il and i2, and a time specification indicated by the $\langle at \rangle$ relation.

The generation process takes such representations as input, and produces sentences using augmented phrase structure rules, looking up the specific lexicon and grammar of the target language.

If sentences are generated in several languages from the same language independent semantic representation, they can be considered as paraphrases and should thus be regarded as mutual translations.

Such a system offers advantages in that translations are obtained without the multiplication of modules necessary with transfer-based methods.

In the representation above, i1 and i2 are individual instances:

- a) They are actual occurrences of types like *girl* and *boy*;
- b) In contrast to event concepts (like *go*, *eat*, etc.) or quality concepts (as *big*, *white*, etc.), individual instances presuppose a stable referential category: in other words, they refer to entities in the real world (Kleiber 1981: 39-67).
- c) In addition, instances are defined by a relationship of quantity <quant> and a relation-ship of determination <det>. Roughly speaking, the first defines the number of considered elements, and the second indicates a relation to the universe of discourse. For example, the value DEF indicates a situation where the knowledge of the instance is already shared by the speakers.

Individual instances are most often linguistically coded into NPs. It has to be noticed that:

- a) generic expressions are not individual instances: indeed, there is neither semantic quantification (generics do not express a set of particular elements) nor relationship to a specific universe of discourse. Moreover, generic expressions refer to types, either directly or by means of an operator (Carlson 1978, Krifka 1987, Heyer 1990), but they do not denote an occurrence.
- b) events or qualities can be nominalized: the result is then a noun phrase that is not an individual instance, as in the following examples:
 - (1) The victory of Caesar The kindness of Caesar

INDEF.

or in the nominal realization of events such as (Galmiche 1983):

(2) Le piano est difficile La marche à pied est bonne pour la santé

or again in noun phrases in verbal collocational expressions:

(3) faire une promenade (= se promener) faire une visite (= visiter)

DESCRIPTION OF THE QUANTIFICATION AND DETERMINATION The determination relation

As far as individual instances are concerned, the relationship to the universe of discourse is expressed by the relation <det>, which can take either of two values: DEF or

DEF indicates that a referent (or a referential set) satisfying the sorting information provided by the concept has been previously mentioned in the discourse situation (Hawkins 1978, Heim 1982).

INDEF is an abstract discourse operator that introduces an entity for the first time in the discourse situation (Seuren 1985, Heim 1982, Kamp 1981).

Operators can also apply on DEF/INDEF to indicate more precisely how the referent is introduced: an operator *DISTR* applied to *DET* picks up the individuals in the situation one by one to give rise, in combination with the information conveyed by the quantification, to distributive expressions like *every x*, *chaque x*.

An operator *NOTSPEC* (Fodor and Sag 1982) can be applied to *INDEF* to select an individual at random and produce, for example:

(4) You can buy any book

On the contrary, the operator SPEC marks that the speaker is introducing a new item in the universe of discourse, but that the item is known by itself:

(5) J'ai vu un certain étudiant

The relationship <det> can point to particular sections of the universe of discourse. For example, if an object is bound to the speaker, the instance will be realized by a personal determiner (my, mon, and so on). The representation:

(6)

```
i1 -isa> *wife*
-points_to> P1
```

where <points_to> indicates a relation with a particular area of the universe of discourse, and P1 symbolises the speaker, will give:

(7) Ma femme My wife

The addressee is represented by P2, and the referent by P3. These can be merged to produce the determiners *notre*, *our*, etc.

Demonstratives will be generated much in the same way, but instead of pointing to one of the interlocutors, they will be indicators of the context (S0). The representation

(8)

```
i1 -isa> *apple*
-points_to> S0
```

is the source of the noun phrase:

(9) Cette pomme This apple

The quantification relation

The quantification relation <quant> takes one of the basic values SING or PLUR. This relation builds a set: SING denotes a set with at most one element of the considered type whereas PLUR constructs a set with several elements of the same type.

These basic values can be modified by values such as:

```
(10) a) {0, 1, 2, 3,..., n}
b) {ALL, FEW, MANY, MOST, COLL}
```

The elements of (10) are defined as relations between sets, after the generalized quantifier approach (Barwise and Cooper 1981, Westerstähl 1986). In the sentence

(11) All the children are sleeping

All indicates that the set A of children is included in the set B of the sleeping ones. The definition of (10) is taken as:

```
a) 0(A,B) iff A ∩ B = 0

1(A,B) iff |A ∩ B| = 1

2(A,B) iff |A ∩ B| = 2

n(A,B) iff |A ∩ B| = n

b) ALL(A,B) iff A ⊆ B

MOST(A,B) iff |A ∩ B| > |A-B|

MANY(A,B) iff |A ∩ B| > m

FEW(A,B) iff |A ∩ B| < m (m found from context)

COLL(A,B) iff A={x|P(x)} & MOST(A,B)
```

The COLLection operator builds a set of x considered as a set (collective reading). This operator is often lexicalized:

```
(12) Un bouquet de fleurs
Un troupeau de vaches
Une chaîne de montagnes
```

Other operators can also be applied to the selected quantifier: for example, the MONOTONE+ or the MONOTONE- operators (Barwise and Cooper 1981) define an orientation scale: MONOTONE+ places a lower bound, and conversely a MONOTONE-places a upper bound: applied to FEW, MONOTONE+ gives plusieurs, several, etc., whereas MONOTONE- applied to the same will give quelques, a few, etc.

Scope

It is also possible to account for the scope of the quantifiers: the scope is indicated in the representation whenever the order of the arguments is not the natural order stemming from the conceptual structure. The sentence:

(13) All women love a man

has different readings:

```
(14)

(a) \forall x (\exists y (W(x)&M(y) \longrightarrow L(x,y)))

(b) \exists y (\forall x (W(x)&M(y) \longrightarrow L(x,y)))
```

The first reading (a) has the representation:

where i1 takes i2 in its scope, whereas reading (b) will be represented as:

where i1 is placed in the scope of i2.

Generation

The generation then proceeds along the steps below:

- a) -isa> yields the head of the constituent. An appropriate lexeme for the specific language is then assigned by looking up the corresponding lexicon: in this case, *child* is replaced by child in English, or enfant in French.
- b) a structure such as -quant> ALL(PLUR) -det> DEF

will be generated by unification with some specific determiner of the target language realizing the features [ALL(PLUR), DEF].

- c) finally, agreement operates to obtain:
 - (15) All the children Tous les enfants

MASS EXPRESSIONS

Definition of mass terms

Mass expressions can be distinguished by syntactic and semantic tests. Syntactically, cardinal numerals modify count nouns, never mass nouns. In many languages, certain determiners are compatible only with one class or another. For example, in English, *little* and *much* modify mass nouns, *few* and *many* modify count nouns. The same type of opposition (*un peu de, peu de*) can be found in French, where the *partitive article* (*du, de la*), which accompanies the mass nouns, also exists.

Semantically, two criteria are commonly proposed: the cumulativity and divisivity of reference (Quine 1960). Mass nouns get the feature +mass, whereas count nouns are given +count.

The cumulativity criterion expresses the fact that some water (+mass) added to some water is still some water (and not two waters), whereas a dog and a dog make two dogs.

The criterion of divisivity says that any part of water always remains some water.

These properties can be explained by the view that mass nouns describe homogeneous masses (Bunt 1985, ter Meulen 1981). This is not the case for the referent of a count term which is inhomogeneous in the sense that some parts of it are more important than others, and it is not the case that any part can be removed without a change. A dog without a tail is still a dog, but a dog without a head is not a dog any more.

Abstract terms such as *passion*, *love*, etc. are often considered as mass expressions when they are used in the singular: the determiner is indeed the same as that for prototypical concrete mass terms as shown in (16):

```
(16) J' ai un peu de sympathie pour vous de la
J' ai un peu de bière de la
```

Such abstract terms comply with the semantic criteria of cumulativity and divisivity.

In the plural, there is a change of category, abstract mass terms becoming count terms, just like concrete terms.

The puzzle of mass expressions

Pelletier and Schubert state at the beginning of their study of mass expressions that the problem of mass expressions provides a rich area in which to formulate and test various logico-linguistic theories" (Pelletier and Schubert 1989: 327).

Indeed, these expressions pose an essential problem for theoretical approaches that give an important place to a conceptual component. Some expressions can be +mass or +count (17), others are mostly +count(18), while others are essentially +mass(19):

- (17) (a) How many whiskies did you drink?
 - (b) How much whisky have you drunk?
- (18) (a) Le professeur est gentil
 - (b) * manger du professeur
- (19) (a) I spilt some water
 - (b) * I spilt one water

Several theories have been proposed to account for these phenomena. We briefly summarize the different proposals below.

200 Meta, XXXIX, 1, 1994

The expressions bear an inherent semantic mass/count feature

If it is taken for granted that the expressions must receive semantic mass/count features from the ontology, several solutions are possible, among others:

- a) The various readings are explained by homonymy (McCawley 1975): for example, there are two lexemes *whisky* corresponding to two different entities, the first one +count, and the other +mass.
- b) All the expressions have a basic semantic mass/count feature: when an expression has two readings, the basic feature is +count, which implies the existence of a mechanism to generate mass terms. This solution was proposed by Pelletier (Pelletier 1975), who suggested a universal grinder for mapping count nouns into mass nouns.
- c) Another solution is similar to the previous one, except that the basic feature is +mass. Bunt (Bunt 1985) has proposed such a solution, with a universal sorter to produce the count nouns.

All these proposals have been questioned:

The view (a) above that assumes homonymy of expressions is plausible for sentences such as (Gillon 1992):

(20) How many chickens are in the yard? How much chicken should be served to each guest?

As a matter of fact, the relationship between the chicken on my plate and the animal in the farm yard is not immediate. I can match the two referents because I know, by encyclopaedic knowledge, that there is a relationship, but the two objects have very different properties. With a naive look at the living animal and the cooked meat, what do they still have in common? They are regarded as different entities but both are called by the same name *chicken*: thus there is homonymy.

Besides, there is sometimes a different lexicalization in such cases, e.g., mutton or sheep, which shows that common sense considers such objects as unbound.

However, in sentences (17) the two occurrences of *whisky* have the same properties: the expression *whisky* obviously refers to the very same entity, and there is no homonymy.

Proposals (b) and (c) set up an ontological problem (Kleiber 1990): an object does not change its fundamental properties according to whether it occurs as a mass or a count noun.

Red wine is still red wine in the noun phrases

(21) some wine a glass of wine

To give another example, in the following sentence

(22) Il vend de l'oeuf

the seller supplies individual eggs, and not a pulp made of eggs.

The hypothesis of a *universal grinder* or the *universal sorter* transforming the nature of the entities remains very questionable.

Moreover, changing features by such devices can lead to unexplained results. A count term transformed into a mass term produces a generic reading

(23) faire du veau (= élever des veaux)

which is hardly explained by the model without additional intricacies.

Other mechanisms such as a *universal packager* have been suggested to transfer mass to count terms comprising a measure:

```
(24) some wine a bottle of wine a glass of wine
```

However, the multiplication of devices changing one category into another can be considered as a very *ad hoc* mechanism.

The apparent failure of theories proposing an inherent mass/count semantic feature leads us to another, more radical, position.

The expressions do not bear inherent semantic mass/count features

The expressions do not bear features such as +mass or +count. This point of view is also expressed by Damourette and Pichon, quoted by Galmiche (Galmiche 1988: 188): "toute substance peut être envisagée d'un point de vue ou de l'autre".

This can be done in various ways. Two are analysed below.

a. The determiner produces the mass/count reading

The interpretation depends on the determiner that should convey the information about the mass or the count characteristics of the referent (Bunt 1985). In other words, the mass/count feature is not a property of the noun, but of the NP *through* the determiner. The process of transforming count into mass terms is not due to the noun.

Unfortunately, this solution is not quite satisfactory.

Let's consider the following examples:

(25) a. Beaucoup de personnes b. Beaucoup de fumée

It is intuitively clear that (a) is a count expression whereas (b) is a mass expression, though the two determiners are identical. To maintain the claim that the quantifier determines the mass or the count properties of an expression, it must be postulated that *beaucoup* de is a homonym in French, now a mass operator, then a count operator.

Moreover, it remains to be explained why certain expressions cannot take the +mass feature, as in:

```
(26) * spill some teacher
```

Conversely, it also remains to be explained why, without an appropriate context, such as contrast or enumeration, it can be difficult to build a count NP from certain mass expressions (Galmiche 1988):

```
(27) il y a deux intelligences: celle du coeur et celle de l'esprit.
?? il y a deux intelligences
```

To explain these observations, one must call for an inherent feature which blocks NPs like (26) and (27).

b. The semantic relation to the conceptual type rules the mass/count feature
Another interesting proposal was made by Lerot in a more semantic perspective
(Lerot 1992). In his view, an instance can be related to a concept type by two relations:

```
(28)
(a) (un) sanglier — isa> - *boar*
(b) (du) sanglier — <portion> - > *boar*
```

202 Meta, XXXIX, 1, 1994

In (28a) the relation rules count expression, whereas in (28b), a mass expression is selected.

In both cases, the operation takes place at the semantic level and is thus language independent, a valuable asset for multilingual generation.

However, this proposal has the same drawbacks as the one above, as examples such as (26) and (27) have to be blocked, which would require a mass/count semantic feature to be attached to the concept type.

To sum up

Postulating inherent mass/count features or rejecting them seems equally unsatisfactory. However, some theory has to be proposed to account for the mass/count puzzle in generation.

A tentative solution

Kleiber (Kleiber 1990: 82), relying on (Weinrich 1966) and (Allan 1980), formulated a way of dealing with the difficulties of mass/count expressions. He suggests accounting for them by considering two levels:

- a) an internal level which denotes an inherent property of the considered entity;
- b) an external level, that of the NP, where the determiner plays a prominent role: it determines the mass/count feature of the NP as a whole.

(29):

In contrast with the *universal grinder/sorter* approach, the entities keep their inherent features. This makes it possible to explain that some whisky is always some whisky, while at the same time two whiskies are two instances that must be in some form of packaging (bottle, glass, barrel) to be picked up, depending on the pragmatic context.

However this proposal is still not absolutely satisfactory.

- a) At the semantic level the determiner is not yet available. In the multilingual generation perspective, it is therefore necessary to analyse the determiner in terms of mass/count value.
- b) Transforming mass terms into count terms or conversely is handled in much the same way. This is difficult to accept, as a count term transformed into a mass term is always very strongly marked, as shown by examples:

(30) bouffer du curé(31) vendre de la femme

Moreover, there can be an aspectual interpretation of habit or intensity resulting from the transformation of count terms into mass terms.

It is thus necessary to revise the Kleiber hypothesis. We therefore propose to keep the idea of the double level of structuring, but remove the determiner and replace it by a semantic feature: (32):

count term

PLUR +

Count NP

mass term

where PLUR operates on an individual instance that has the inherent feature +mass or +count.

When the argument is a count term, the result is always a count term, but when the argument is a mass term, a change of category is produced: the mass term becomes a count term. This requires a packaging unit:

(33) One glass

of whisky

Two bottles

of wine

Three barrels

of beer

The measure is often not expressed because it can be inferred from the knowledge of the world: at the bar, one can order a *glass of beer*, not *a barrel*. It is thus pointless to mention the unit explicitly. If the measure is indicated, the representation is:

```
il - isa> *whisky*
```

- quant> 2(PLUR) -measure> *glass*
- det> INDEF

The second part of the rule proposed by Kleiber (b) is not related to individual instances but, as observed above, count terms used in mass NPs in a predicate indicate a mere aspectual change, or a quantification over the predicate: for example, in French, the predicate containing a partitive article + count term indicates a habit or an intensity aspect:

- (34) vendre de la femme
- (35) faire de la vitesse

To account for the individual instances, rule (32) is sufficient. This hypothesis offers the following benefits:

- the notion of PLUR is of a semantic nature; it is independent of specific languages, and can be used in semantic representations for multilingual generation;
- b) the different behaviour between category changes for mass terms and count terms can now be accounted for:
- c) in this view the desired blocking of such phrases as:
 - (36) * manger du professeur

is realized, according to the intuition.

Mass expressions are all generated from feature structures like:

(37) INDEF + SING(+mass concept)

which will yield a mass reading, with e.g. a partitive determiner.

The other cases produce count readings, as in the special case:

(38) INDEF + PLUR(+mass concept)

204 Meta, XXXIX, 1, 1994

It can also be noted that quantifiers can be worked into these rules as usual:

(39) INDEF + MOST(SING(+mass concept)

which would yield for example beaucoup d'eau.

Remaining problems

This solution entails that in the ontology concepts are marked with a +mass or +count feature: this is a rather plausible hypothesis in naive semantics, where the objects are categorized according to their cognitive properties in the real world in an ontology organized with attribute-value pairs and inheritance rules.

However a certain disparity between languages may exist, and a same reality can be expressed as mass in one language and as a count term in another: for example, one can find such a case between French and English (Pelletier and Schubert 1989):

mass term	count term
dandruff (UK) furniture (UK) vaisselle (FR)	pellicules (FR) meuble (FR) dishes (UK)

Solutions are being sought in the development of a conceptual grammar. The generation process can nevertheless be carried out, as such cases are well described in contrastive grammars: they can thus be handled in a practical way, pending the full development of a more comprehensive theory.

CONCLUSION

Multilingual sentence generation in the GENESE project was presented. The attention was mainly focused on generating individual instances, including mass/count expressions.

It has been shown how noun phrases are generated from semantic representations using the relationships of quantification and determination, with the help of a number of operators.

The hypothesis has been put forward that concepts in the ontology have mass/count features, obtained by inheritance. It has been shown that mass expressions could be generated from a feature structure containing only a few features.

A number of problems remain for further research, such as difficulties concerning mass/count variations in different languages.

Note

1. Project funded by the Communauté Française de Belgique under contract number 89/94-137.

REFERENCES

ALLAN, K. (1980): "Nouns and Countability", Language, pp. 541-567.

BARWISE J. and R. COOPER (1981): "Generalized Quantifiers and natural Language", Linguistics and Philosophy, 4, pp. 159-219.

BUNT, H. (1985): Mass terms and Model-Theoretic Semantics, Cambridge, Cambridge University Press.

CARLSON, G. N. (1978): Reference to Kind in English, repr. 1980, New York, Garland.

FODOR, J. D. and I. SAG (1982): "Referential and Quantificational Indefinites", Linguistics and Philosophy, 5, pp. 355-398.

GALMICHE, M. (1983): "L'utilisation des articles génériques comme mode de donation de la vérité", LINX, 9, pp. 29-83.

GALMICHE, M. (1987): "A propos de la distinction massif/comptable", Modèles Linguistiques, IX,2.

GILLON, B. S. (1992): "Towards a Common Semantics for English Count and Mass Nouns", *Linguistics and Philosophy*, 15, pp. 597-632.

HAWKINS, J. A. (1978): Definiteness and Indefiniteness: A study in Reference and Grammaticality Prediction, London, Croom Helm.

HEIM, I. (1982): The Semantics of Definite and Indefinite Noun Phrases, Doctoral Dissertation, Univ. of Massachusetts, Amherst.

HEYER, G. (1990): "Semantics and Knowledge Representation in the Analysis of Generic Descriptions", Journal of Semantics, 7, pp. 93-110.

KAMP, H. (1981): "A Theory of Truth and Semantic Representation", Groenendijk, J. A., Janssen, T. M. and M. B. Stokhof (Eds.), Formal Methods in the Study of Language, 277-322, Amsterdam, Mathematical Centre Tracts, pp. 135.

KLEIBER, G. (1981): Problèmes de référence : descriptions définies et noms propres, Paris, Klincksieck.

KLEIBER, G. (1990): L'article LE générique. La généricité sur le mode massif, Genève, Droz.

KRIFKA (1987): An Outline of Genericity, SNS-Bericht 87-25, Tübingen.

LEROT, J. (1992): Vade-mecum pour la représentation de contenus assertionnels, UCL, GENESE Research Group, Internal working document.

McCAWLEY, J. (1975): "Lexicography and the Count-mass Distinction", Proceedings of the First Annual Meeting of Berkeley Linguistic Society, pp. 314-321.

PELLETIER, F. J. (1975): "Non-singular Reference: Some Preliminaries", Philosophia, 5.

PELLETIER, F. J. and L. K. SCHÜBERT (1989): "Mass Expressions", Gabbay, D. and F. Guenthner (Eds.), Handbook of Philosophical Logic, Vol. IV, Reidel Publishing Company.

QUINE, W.v.O. (1971): "The Inscrutability of Reference", Steinberg, D. and L. A. Jakobovits (Eds.), Semantics: an Interdisciplanary Reader in Philosophy, Linguistics, and Psychology, 142-156, Cambridge, Cambridge University Press.

SEUREN, P. A. (1985): Discourse Semantics, Oxford, Basil Blackwell.

WEINREICH, U. (1966): "Explorations in Semantic Theory", Current Trends in Linguistics, Vol III, The Hague, Mouton, pp. 395-475.

WESTERSTAHL, D. (1989): "Quantifiers in Formal and Natural Languages", Gabbay, D. and F. Guenthner (Eds.), *Handbook of Philosophical Logic*, IV, 1-131, Reidel Publishing Company.

WESTERSTAHL, D. (1989): "Quantifiers in Formal and Natural Languages", Gabbay, D. and F. Guenthner (Eds.), *Handbook of Philosophical Logic*, IV, 1-131, Reidel Publishing Company.

WESTERSTAHL, D. (1989): "Quantifiers in Formal and Natural Languages", Gabbay, D. and F. Guenthner (Eds.), *Handbook of Philosophical Logic*, IV, 1-131, Reidel Publishing Company.