

Universal Quantification in Machine Translation¹

Cornelia ZELINSKY-WIBBELT

IAI / EUROTRA - D

Martin-Luther-Straße 14

D-6600 Saarbrücken 3

Abstract

This approach has been developed in the context of the EUROTRA machine translation (MT) project and thus has been designed with respect to a syntax based stratificational translation process.² We assume that in a semantic representation determiners are deleted and that their semantic function which is represented by semantic features is percolated into the mothernode. The semantic functions of determiners are explicated. The interaction between grammatical and lexical quantification is outlined. Ensemble theory is applied to the "count"/"mass" noun distinction. Transfer of quantification between German, English, and French is illustrated with respect to the "count"/"mass" distinction. The article closes with an outlook on the relevance of generalized quantifiers for Machine Translation.

1. Semantic representation of determiners in EUROTRA

EUROTRA aims at defining a semantic representation which guarantees simple transfer between all European languages, that is, it should be "euroversal". The concept of "euroversality" implies, amongst others, a semantic representation in a canonical form out of which all European languages may be generated. With respect to this canonical form it is reasonable to delete the determiners during translation into the semantic representation and to represent their meaning by semantic features of the NP. This step may be motivated primarily by two facts:

- (1) Languages vary with respect to the use of 0-determiners.
- (2) The set properties realized by an entity are expressed differently in different languages.

The idea that determiners are not deep structure constituents, but that their surface structure constituents have to be generated from a semantic representation is not new. It can already be found for example in PERLMUTTER 1970. Moreover, to represent the quantifier by means of semantic features of the NP implies that the entity which is focussed by the process of determination cannot be referred to directly, but only as the argument of the determiner which provides a new referent (cf. PINKAL 1986). BARWISE & COOPER (1981) consider determiners as two-place predicates which take the noun which is the domain of quantification as one argument, and the rest of the sentence, which is the predicate quantified as the other argument. With respect to the EUROTRA MT system this has important implications for the translation between the syntactic dependency level - the EUROTRA Relational Structure (ERS) and the semantic level - the Interface Structure (IS). Determiners which have the function of modifying nouns at ERS on the basis of several syntactic conditions establish different types of determination. Those types of determination are the basis for deducing (i.e. translating) exactly that information which yields the new referent in the NP by unifying with the semantic features of the noun.

Although both determiners and quantifiers have characteristic functions, they have others in common, so that a borderline is difficult to draw. Cases of crossclassification exist in many languages, as for example the one constituted by German *der/ein/0*, French *le/un/0*, English *the/a/0* (cf. VATER 1963). This is why we describe both determiners and quantifiers by a common set of semantic features.

2. The semantic functions

It is agreed in the literature that determiners and quantifiers share the function of DELIMITATION (cf. VATER 1980). This delimitation consists in the localisation of a referent in the speech or textual context or the non-linguistic situation or in relation to the presupposed knowledge of the hearer or reader (only the first of these functions, and this again in a rather restricted way, may be represented in the EUROTRA system). BARWISE & COOPER (1981) refer to this function of delimitation as the property "lives on" and define that determiners "assign to common count noun denotations (i.e. sets) A a quantifier that lives on A." (BARWISE & COOPER 1981.179)

2.1. Quantification over whole sets: "generic" versus "identifying"

It is the function of determiners and quantifiers to **quantify over sets of entities**. The writer's motivation to create sets is that the entities which should be members of the set share one or several properties. Following the tradition of the MONTAGUE approach, BARWISE & COOPER treat all NPs as quantifiers which denote sets of properties of individuals. There are two basic types of WHOLE SETS, which may be created.

- (1) The entity's **extension** is created "generically" by means of its inherent lexical meaning as in the following example:

Die Linguisten sind in formalen Sprachen geübt
(Linguists are practised in formal languages.).

Here the NP quantifies exactly over the complete set of linguists of the actual world.

- (2) An **Intensional** property of the entity set makes possible its "identification". In this case a **WHOLE SET** is referred to which is precisely delimited (cf. VATER 1963, PLATTEAU 1980). This type of entity set may only be established context-sensitive. It is thus a set which may be referred to as a **WHOLE SET** only with respect to a certain domain of interpretation, which is the intensional property:

The linguists of EUROTRA ...

This NP quantifies exactly over that set of linguists who work for EUROTRA.

2.2. The semantic functions of determiners: determiners as variables and as variable-binding functions

It is the function of determiners to **select one or several entities from a set of entities** (cf. PLATTEAU 1980).

The salient function of indefinite determiners is equivalent to that of the existential quantifier (cf. LANGENDONCK 1980, PLATTEAU 1980); they introduce new entities into the speech or text situation. Thus they only express that entities exist in the speech situation, without "specifying" which. It is an **infinite set of a potential of entities** (cf. HAWKINS 1978.198). We may therefore say that indefinite determiners in their salient function are variables. This yields a **PARTIAL SET** of entities which is "existential".

Beside this salient function the indefinite determiner may also "specify" entities, if it is clear in the universe of discourse which entity is designated (cf. OOMEN 1977 and DI EUGENIO 1986).

The salient function of definite determiners is that the existence of an entity is already presupposed by the writer, i.e. the writer presupposes that the entity is already given, that the reader is already acquainted with it (cf. OOMEN 1977). Now the variable which is presupposed to have been assigned to the entity (entities) by the indefinite determiner is bound:

We need another linguist for EUROTRA-D. The linguist should be a specialist in syntax.

The definite article thus yields a WHOLE SET of entities which is precisely limited by a fixed reference point, that is it is "identified".

2.3. Classifiers

A special case of indefiniteness may be said to be what LANGENDONCK calls "indefiniteness with asserted partition" in opposition to "ordinary asserted partition". We said above that it is the function of determination to select an entity or entities out of a set. We can also say that they partition a set into those entities which are members of a subset and those which are not. With "ordinary asserted partition" expressed by an indefinite determiner and a noun this partitioned set is an infinite set of a potential of individuals. "Indefinites with asserted partition" are classifier constructions. They constitute the clearest instance of exclusiveness or partitioning, in particular partitive constructions with the semantic structure

$$\exists x^1 (x^1 \subseteq x^2).$$

(cf. LANGENDONCK 1980.213). Exactly the same holds for this structure that holds for the relation between definite and indefinite determination: A potential subset of entities is presupposed, when a specific part of those is asserted:

this part of the article ...

Beside the feature "partitive" the features "sortal", "collective", "mensural", "scale", and "numerative" become relevant in the realization of this structure.

3. Determination of the set properties

The fact that the set to be quantified is greater than one is expressed by different surface structures in the European languages. With proper "count" nouns plurality may be designated by the plural morpheme (*the determiners*). With "discontinuous" "mass" nouns classifiers may be used in order to partition the mass into amounts and thus make the partitioned masses (not the mass on its own!) countable (*several pieces of advice, different boxes of vegetables*). Finally, a "collective" refers to a set which is greater than one (*the furniture*). In German the individuation of certain "abstract" "mass" entities may simply be achieved by the plural morpheme. The use of the German plural is only impossible with nouns which designate "continuous" "masses".

3.1. The interaction of lexical and grammatical quantification

A noun designates an entity the inherent setforming properties of which are lexicalized. By means of grammatical quantification this entity may form different sets. On the one hand there are entities, the inherent setforming properties of which may not be influenced grammatically, but which may only designate on their own. This is the case with "continuous" "mass" nouns; we may also say that they designate sets absolutely. On the other hand there are entities which are not able to form sets on their own. This holds for "discontinuous" "mass" entities; they may also be considered as designating sets by a variable with respect to their lexical potential, this variable only being filled by a constant by grammatical context. From a logical point of view this idea is developed more precisely and moreover integrated into a coherent system in BUNT 1979 and 1985. In the following we will apply this system to language.

"Continuous" ensembles (cf. BUNT) are true "masses" which may not be enumerated that is they may not be designated by a plural expression. They satisfy QUINES cumulative reference

condition, or more precisely, the distributive reference condition. The cumulative reference of mass nouns implies that the union of any two masses W is again W. Or vice versa the distributive reference condition means that any part of some mass W must again be W. If we refer to "continuous" ensembles, we do not imagine any smallest part of the ensemble which may not be divided any more without the ensemble ceasing to be what it was. A prototypical "continuous" ensemble is that referred to by the "mass" noun *time*. The following syntactic condition holds:

- (1) All nonpluralizable "mass" nouns are "continuous".

This means that the property of continuity is lexical. Examples of nouns referring to "continuous" "mass" entities are *participation, impetus, increase, adhesion, importance, extent*.

Contrary to the mode of reference to "continuous" masses is that to "discontinuous" ensembles or sets. While the feature "discontinuous" is lexical, its subspecifications are only realized in interaction with grammatical structure:

"Atomic" sets or ensembles cannot be imagined to have any genuine (=nonempty) parts, that is

- (2) All "count" and "discontinuous" "mass" entities designated by singular nouns are "atomic".

"Atomic" sets or ensembles may, however, be merged into "discrete" sets or ensembles which are constituted either by "individual" "count" entities or by entities which are basically "mass", but which may be turned into an ensemble which we conceive of as having genuine parts e.g. by being represented by several amounts. This is expressed by pluralization or by preceding classifiers, as e.g. with *advice*, which gets enumerable by the "numerative" *piece*. This is not possible with "continuous" mass entities, as e.g. those designated by the nouns *importance, research*.

Moreover, "collectives" are "discrete". Now we can summarize:

- (3) The designation of "discreteness" is yielded by pluralized "count" and "mass" nouns as well as by "collective" nouns (cf. ALLAN 1976.99, where he defines the result of collectivizing as the unmarked (singular) form of plural reference).

4. Transfer of quantified nounphrases

We start from our condition developed in the previous chapter that pluralizability is represented by the lexical features "discontinuous" and "continuous". Singular NPs then have to be translated into three semantically different NPs at IS:

- (1) Into an "atomic" NP if and only if a "count" or "mass" noun for which "complexity" does not equal "collective", and for which "distribution" equals "discontinuous".
- (2) Into a "continuous" NP if and only if a "continuous" "mass" noun is generated.
- (3) Into a "discrete" NP then and only then, if a "collective" noun is generated.

The source IS-representation of an atomic NP will be transferred into the identical target IS-representation with the exclusion of the features "mass" and "count", which may change as in the translation from *le conseil* in it's "individual" reading to English *the advice*, as illustrated in figure 2.

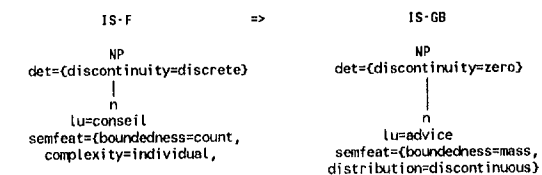


Fig. 1 Transfer from *le conseil* to *the advice*

In this case a singular NP will be generated in the English synthesis.

A "continuous" NP may change into a "discontinuous" "atomic" or "discrete" NP, as in the translation from *der Rat* into *the advice*, as represented in figure 2.

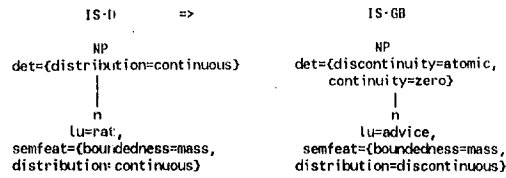


Fig. 2 Transfer from *der Rat* to *the advice*

Again a singular NP will be generated in English synthesis.

NPs referring to "discrete" "mass" entities may either change into a "discrete" NP constituted of "individual" entities, as in the translation from *the furniture* to *die Möbelstücke* or they may be transferred into the same target-language representation by translating into *die Möbel*. The translations are represented in figure 3. Both representations will effect the generation of a plural NP in German synthesis.

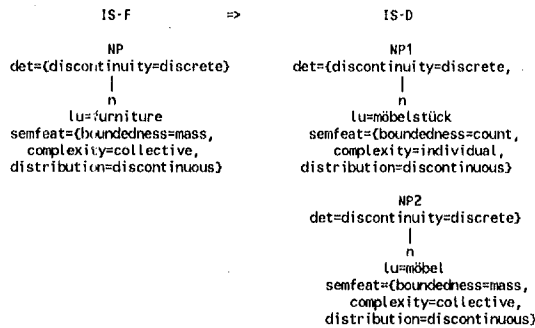


Fig. 3 The translation of *the furniture* into German

During analysis plural NPs are dealt with very simply: they are all translated into discrete NPs at IS. In the same way as with singular NPs, the set properties may change in transfer as in the translation from plural *les conseils* in its collective as well as in its individual reading to singular *der Rat* as represented in figure 4.

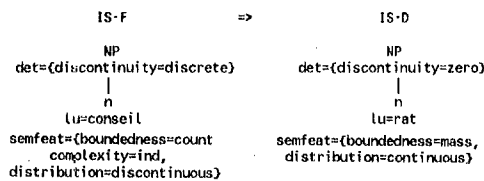
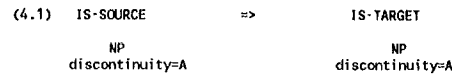


Fig. 4 Transfer from *les conseils* into *der Rat*

The feature "continuous" blocks pluralization. In the case of numeral quantification the German noun must be DISCONTINUOUS; in this case unification succeeds with

Ratschlag, which is "count", "individual", "discontinuous", that is "atomic" in the case of a one-element set and "discrete" in the case of a set that has more than one element. Whereas the latter case is the unmarked case in which the default rule (4.1) applies, the former case is the marked one which is represented in figure 4.



In transfer *les conseils* which is lexically "count" in one reading, "mass" in the other and "discontinuous" in both readings, the latter feature being grammatically specified as "discrete" goes to *the advice*, which has the lexical features "mass" and "discontinuous" the latter feature being subspecified as "discrete" by our default rule (see above rule (4.1) chapter 4.1.), (1) because it is enumerable by means of the numerative *piece* (2) because we may refer to a single representative of the entity in its "atomic" meaning and to a set of representations in its "discrete" meaning.⁴ This translation is represented in fig. 5.

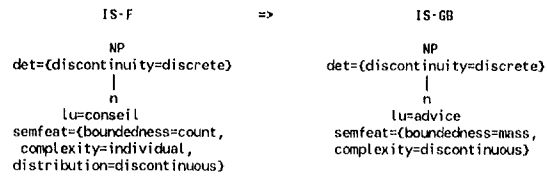
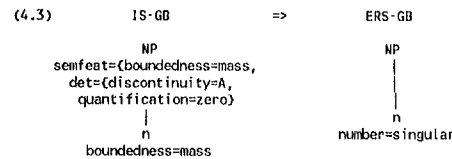
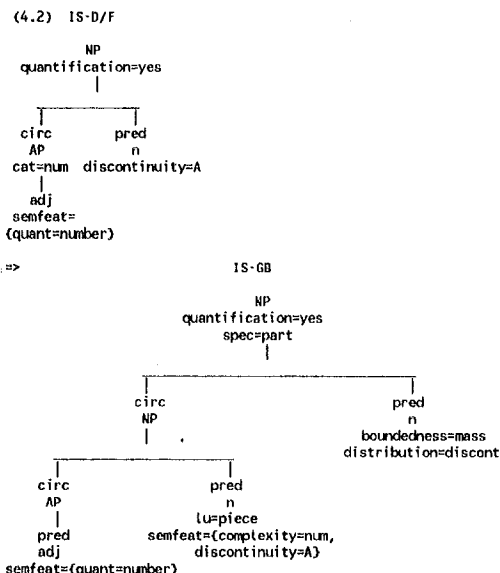


Fig. 5 Translation from *les conseils* to *the advice*

Now rule (4.3) should guarantee for English generation that "atomic" and "discrete" "masses" are translated into a nonpluralized noun in English if the English noun is semantically "mass" and is not modified by a quantifier.



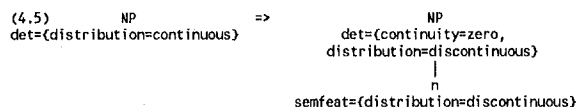
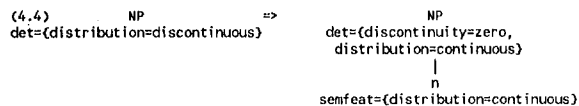
Rule (4.2) guarantees that "discrete" or "atomic" "masses" which are preceded by a quantifier, are translated into a noun which is syntactically governed by the numerative *piece*, which then in turn will be the bearer of the respective singular or plural morpheme which is deduced from the semantic features "atomic"/"discrete".



A sentence-based interpretation will yield *the advice* in analysis as ambiguous between the "atomic" and the "discrete" reading (the "discrete" reading again between the "identifying" and the

"generic" reading) and with some exceptions we must accept it as the correct result of a sentence based analysis to get two translational results in French as in this case: *le conseil* and *les conseils*.

The unmarked transfer is achieved by our default rule (4.1). The marked case is that the lexical value of the target language disagrees with the source language one, so that for the latter case we have rules (4.4) and (4.5).



Moreover by the given transfer rules translations between the following representations will be guaranteed:

- (1) *Les meubles* which is lexically either "count" and "individual", and hence "discontinuous", or "mass" and "collective", that is, it is also "discontinuous" in this second reading. On the basis of their morphosyntactic behaviour both readings yield a "discrete" NP
- (2) The translation from French into English yields two identical translations, as only one lexical unit with the "collective" reading exists in English: *the furniture* which is lexically "mass", "collective", and hence yields a "discrete" NP, so that in the case of being quantified the quantifier is again followed by *piece*. One of the identical readings has to be killed.
- (3) If we translate from French into German, the NP with the "individual" noun is translated into *die Möbelstücke*, which has the same features, both in the NP and in the n. The French NP with the "collective" n is translated into *die Möbel*, which also has the same features, both for NP and n. In the case of a preceding cardinal number phrase the German noun *Möbelstück* with a morpheme as "numerative" must be generated. The German noun in this case is "count", "individual", "discontinuous", that is "atomic" in the case of a one-element set and "discrete" in the case of a set which has more than one element.

4.1. Conclusion

It was the intention of this chapter to point out how two types of semantic features with lexical and grammatical origin which quantify the nounphrase interact in transfer:

From the dictionaries we generate those semantic features quantifying the set of entities which refer to the constitution of the entity ("count"/"mass", "individual"/"collective"/"partitive"/"sortal", "continuous"/"discontinuous"), while in the unmarked case the setforming properties are transferred from source to target language representation by a default rule ("discrete"/"atomic"). More precisely, an "atomic" set always goes to an "atomic" set, a "discrete" set normally goes to a "discrete" set, it may, however, go to a "continuous" set, if a continuous entity is generated from the dictionary as in the case of the correspondence between *les conseils* and *der Rat/the advice*. In the same way a "continuous" set normally goes to a "continuous" set, it may, however, go to a "discontinuous" set as in the opposite translation from *der Rat* to *the advice*.

5. Generalized quantifiers in Machine Translation⁵

Let us close with an evaluation of the super/subset relationship holding for generalized quantifiers and its relevance for machine translation. Indeed, we are convinced, that the properties of persistency, monotonicity, strength and weakness, conservativity and others which BARWISE & COOPER (1981) and others have introduced are relevant with respect to the disambiguation of

determiner readings and thus have to be part of the semantic representation of the NP. BARWISE & COOPER themselves mention the ambiguity of *a few*, which is monotonously increasing (mon ↑) in its positive reading (*at least a few*) and not monotonous in its negative reading (*only a few*):

- mon ↑: *If (at least) a few linguists implement, then a few linguists work.*
 mon ↓: **If (only) a few linguists work, then (only) a few linguists implement.*
 mon †: **If (only) a few linguists implement, then only a few linguists work.*

The fact that negation reverses monotonicity is realized with mass nouns and pluralized count nouns which in their positive reading appear with zero-article. In the positive reading which is mon ↑ the partitive article is used in French:

- If there is wine that contains 12% alcohol, then there is wine that contains alcohol.*
If there are wine bottles that contain 12% alcohol, then there are wine bottles that contain alcohol.
S'il y a du vin qui contient 12 % d'alcool, il y a du vin qui contient de l'alcool.
S'il y a des bouteilles de vin qui contiennent 12% d'alcool, il y a des bouteilles de vin qui contiennent de l'alcool.

In the negative reading which is mon| simple *de* instead of the partitive article is used in French:

- If there is no wine that contains alcohol, then there is no wine that contains 12% alcohol.*
If there are no bottles of wine that contain alcohol, then there are no bottles of wine that contain 12% alcohol.
S'il n'y a pas de vin qui contient d'alcool, il n'y a pas de vin qui contient 12% de l'alcool.
S'il n'a pas de bouteilles de vin qui contiennent d'alcool, il n'y a pas de bouteilles de vin qui contiennent de l'alcool.

6. The organization of the semantic features of determination

As an overview let us give a graphical representation of the organization of the features. In this representation the ENTITY-node is the axiom and each node is subspecified either by a disjunction of features, which we represent by the solidlined edges, or by a conjunction of features which we indicate by the "+" marked edges.

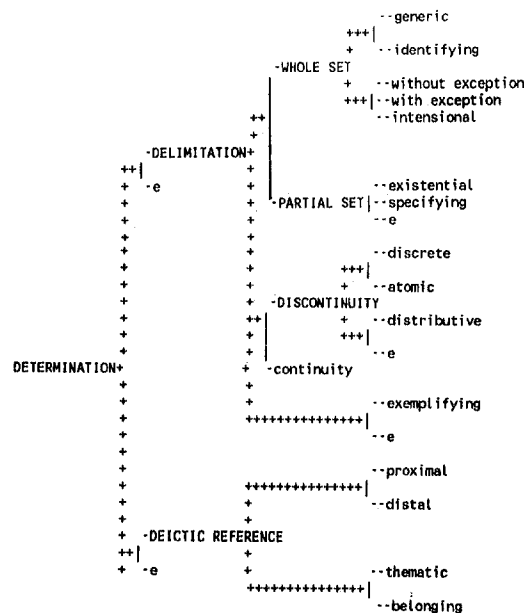


Fig. 11 The organization of the semantic features of determination

7. Summary

We have illustrated several semantic representations which are meant to guarantee the correct generation of different surface structures of quantifiers. The intricate interwovenness between lexical and grammatical quantification has been outlined. In some cases as for example in the case of "collective" and "discrete" or "individual" and "atomic" "discontinuity" the ambiguity could not be resolved.

8. Literature

- ALLAN, K.
1976 Collectivizing. *Archivum Linguisticum* 1.
- ARNOLD, D.J., S. KRAUWER, M. ROSNER, L. des TOMBE, G.B. VARILE
1986 The <C,A>,T Framework in EUROTA: A theoretically committed notation for MT. *COLING*. 1986.
- VAN DER AUWERA, J. (ed.)
1980 *The Semantics of Determiners*. London, Croom Helm.
- BARWISE, J. & R. COOPER
1981 Generalized Quantifiers and Natural Language. *Linguistics and Philosophy* 4. 159-219.
- BUNT, H.C.
1979 Ensembles and the Formal Semantic Properties of Mass Terms. In: PELLETIER, F.J. (ed.) DOWTY, D.R. *Word Meaning and Montague Grammar. The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht.
- BUNT, H.C.
1985 *Mass terms and model theoretic semantics*. Cambridge Studies in Linguistics Vol. 42. Cambridge University Press. Cambridge.
- DI EUGENIO, B., L. LESMO, P. POGLIANO, P. TORASSO, F. URBANO
1986 A logical formalism for the representation of determiners. *COLING* 1986.
- HAWKINS, J.
1978 *Definiteness and indefiniteness: A study of reference and pragmaticality prediction*. London, Croom Helm.
- VAN LANGENDONCK, W.
1980 Indefiniteness, Exemplars and Kinds. In: Van der AUWERA 1980.
- OOMEN, I.
1977 *Determination bei generischen, definiten und indefiniten Beschreibungen im Deutschen*. Tübingen, Niemeyer
- MONTAGUE, R.
1973 The Proper Treatment of Quantification in English. In: J. HINTIKKA, J. MORAVCSIK, and P. SUPPES (eds.). *Approaches to Natural Languages*. Reidel, Dordrecht
- PERLMUTTER, D.M.
1970 On the Article in English. In: BIERWISCH, M. & HEIDOLPH K.E.: *Progress in Linguistics*. Mouton, The Hague.
- PINKAL, M.
1986 Definite Noun Phrases and the Semantics of Discourse. *COLING*. Bonn 1986. 368-373.
- PLATTEAU, F.
1980 Definite and indefinite Generics. In: Van der AUWERA 1980
- QUINE, W.
1960 *Word and Object*. Cambridge, Mass.
- STEINER, E.H., P. SCHMIDT & C. ZELINSKY-WIBBELT.
1988 *New Approaches to Machine Translation*. Frances Pinter, London. Forthcoming.
- VATER, H.
1963 *Das System der Artikelformen im gegenwärtigen Deutsch*. Tübingen, Niemeyer.
- VATER, H.
1980 Quantifier Floating in German. In: Van der AUWERA 1980.
- ZELINSKY-WIBBELT, C.
1986 An Empirical Approach Towards a System of Semantic Features. *COLING*, Bonn 1986.
- ZELINSKY-WIBBELT, C.
1987a *Semantische Merkmale für die automatische Disambiguierung. Ihre Generierung und ihre Verwendung*. EUROTRA-D Working Papers. MARTIN-LUTHER-STR. 14, 6600 SAARBRÜCKEN, TEL. 0681/39313, BTX *(39)921607#
- ZELINSKY-WIBBELT, C.
1988a From Cognitive Grammar to the Generation of Semantic Interpretation in Machine Translation. In: STEINER et al. 1988.
- ZELINSKY-WIBBELT, C.
1988b The transfer of quantifiers in Machine Translation. In: STEINER et al. 1988.
1. A more detailed version of this approach can be found in ZELINSKY-WIBBELT 1988b
 2. For details concerning the EUROTRA formalism cf. D. ARNOLD et al. 1986
 3. For details concerning the semantic features of nouns cf. ZELINSKY-WIBBELT 1988a and 1987
 4. *Advice* is enumerable by the "numerative" *piece* in contrast to other "abstract" English "mass" nouns like *patience, faith, dignity, behaviour, research*. We hypothesize that the "numerative" *piece* in English has a similar function as the plural morpheme has in German with nouns designating "abstract" "mass" entities which are "discontinuous" which means that several exemplars of an atomic entity may be merged into a discrete ensemble (cf. Löbel 1986).
 5. I would like to thank Michael Grabski for discussing this chapter with me.