

Definite marking and referential status in Greek, Swedish and Polish

Ann Lindvall

Introduction

This article presents an analysis of definiteness from a cognitive and typological approach. The claim is that definiteness is a property based on referential grounds, and that this property goes beyond the language specific variations. First, some theories of the concept of definiteness are presented, together with a brief typological overview. Then follows the analysis of three languages: Greek (modern), Swedish and Polish, first with attention to forms and then to stages of referential status. Finally, the results are discussed and formed into a schema, where one can discern typological patterns.

Previous theoretical treatments

Definiteness – definite and indefinite noun phrases – is a complex concept, including properties like referentiality, familiarity, availability, identifiability, inferrability etc. Many have tried to describe the functions of definiteness, both philosophers and linguists. This chapter gives a brief summary of some theories.

The classical approach, with its roots in the logico-philosophical tradition, presupposed reference to The Real World. It saw the question of indefiniteness and definiteness as a binary question and was labelled the *Familiarity theory* by Christopherson 1939, also discussed by Hawkins 1978. Its content is expressed by Heim 1983:93:

“A definite is used to refer to something that is already familiar at the current stage of the conversation. An indefinite is used to introduce a new referent.”

Though this theory has obvious shortcomings, it is useful as a point of departure, as it contains the two key concepts; familiarity and novelty. But it was necessary to extend the concepts outside logic. Karttunen 1976 added

the notion of *discourse* reference. By stressing the importance of discourse, both Karttunen and Heim developed the traditional Familiarity theory.

From a dichotomy of two mutually exclusive categories, new theories describe gradient *scales* of definiteness. Prince 1981, continuing the discourse approach, presents the *Familiarity scale*, here as a brief summary:

- | | |
|-------------------------------------|---|
| a) New: Brand-new: (Un-anchored) | I got on <i>a bus</i> yesterday and the driver was drunk. |
| b) New: Brand-new: Anchored | <i>A guy I work with</i> says he knows your sister. (linked to I) |
| c) New: Unused | <i>Noam Chomsky</i> went to Penn. |
| d) (Non-containing) Inferrable | I got on a bus yesterday and <i>the driver</i> was drunk. |
| e) Containing Inferrable | Hey, <i>one of these eggs</i> is broken! |
| f) (Textually) Evoked | A guy I work with says <i>he</i> knows your sister. |
| g) Situationally Evoked | Pardon, would <i>you</i> have change of a quarter. |

Prince herself points at the difficulties in distinguishing Unused from Inferrable. She also stresses the importance of cultural assumptions required for the inference, a point of view that, in my opinion, cannot be overestimated.

Givón 1984 considers the discourse-scope phenomenon of definiteness and the sentence-scope phenomenon of referentiality. He also stresses the *co-operation* of the speaker and the listener. The background of the definiteness is the universe of discourse, which is negotiated between speakers and hearers. The source from which the hearer assigns reference is the *active discourse file*, which in its turn takes its reference from those (p. 399ff):

- 1) permanent file
 - a) unique physical or cultural entities
 - b) proper names
- 2) immediate deictic context
 - a) absolute deictic availability, e.g. 1st and 2nd person of anaphoric pronouns
 - b) relative deictic availability, e.g. 3rd person of anaphoric pronouns, parts of wholes
- 3) specific discourse file, i.e. the actual information passed from speaker to hearer

Givón, too, stresses the *gradiency* of definiteness. He splits INDEFINITE into REFERENTIAL and NON-REFERENTIAL and suggests the following scale:

DEFINITE > REF-INDEFINITE > NON-REFERENTIAL > GENERIC

One final contribution mentioned here is the label *Givenness hierarchy* by Gundel, Hedberg & Zacharski 1993. This is based on the *cognitive* status of the referent, where each status is implicationally related to the lower:

| | | | | | | | | | | |
|-------------|---|------------------------|---|----------|---|-------------------------------|---|---------------------------|---|----------------------|
| in focus | > | activated | > | familiar | > | uniquely identi- fiable | > | referential indefinite | > | type identifiable |
| it | | that this this N | | that N | | the N | | this N | | a N |

They are relating the cognitive status to the linguistic form, from an indefinite article + noun for the very lowest status, over modifiers like demonstrative determiners / definite article + noun to unstressed or zero pronouns for the very highest status, i.e. in focus. The authors, comparing five languages, find striking similarities between the use of linguistic forms and the degree of cognitive prominence. The results are further related to the Gricean Maxim of Quantity, which gives still more evidence to the hypothesised hierarchy.

Fraurud 1994 discusses the two definitions endophora/exophora and focus maintenance / focus shift, with the conclusion that both need elaboration. As for *indefiniteness*, which has been treated far less than definiteness, a typological discussion on indefinite pronouns was held by Haspelmath 1994. Some earlier analyses of Greek, Polish and Swedish were presented in Lindvall 1994, forthc.

As can be concluded from the binary values of definite/indefinite with reference to some neutralised independent universe, the theories have developed into gradient scales and hierarchies with reference to a discourse universe, negotiated by the two interlocutors. The new descriptions are doubtless helpful both in explaining the universality and the language specific variations of the category definiteness.

Typological overviews

This chapter first presents two typologies, with considerable theoretical differences between them. The first, a basic typological overview of

languages expressing definiteness was done by Krámský 1972. Though with serious shortcomings, this is an impressive attempt to cover all possible types of the coding of definiteness/determinedness. Especially interesting is his comparison of English and Czech, where he stresses the theme/rheme notion as a way of expressing the same semantic and discourse-scope content as the English article. Krámský gives the following suggested typology of the world's languages:

- A. Languages expressing the category of determinedness vs. indeterminateness by means of independent words
- B. One member of the category of determinedness vs. indeterminateness is expressed by an independent word, the other member is proclitic or enclitic
- C. Both (or more) members of the category of determinedness vs. indeterminateness are either enclitic or proclitic
- D. Languages in which the category of determinedness vs. indeterminateness is inherent in the noun itself or in another word category
- E. Languages in which the category of determinedness vs. indeterminateness is expressed by flexion
- F. Determinedness vs. indeterminateness is expressed by stress or intonation
- G. Languages without article

The typology can be criticised because of the overwhelming attention to the form of the article. Also the occurrence of fixed groups is unsatisfying, although Krámský warns for extensive overlapping. One could easily draw the conclusion that the world's languages are *either* 'definiteness languages' or 'non-definiteness languages'. But as can be seen, dichotomisation per se is hardly to be taken seriously any longer, be it the category of definiteness itself or its occurrence in the world's languages.

A basic typology over *constructions*, Greenberg 1978, has the gradient character of the definite article as the core content. He describes the developmental stages of the definite article and discerns three stages. For the purpose of this paper, his 'stage 0' also added. In the following a short summary is given:

- 0 Demonstrative, which develops from being purely deictic to a discourse deictic, identifying an element as previously mentioned
- 1 Definite article, which differs from demonstrative by being compulsory, pointing out an element as 'identified', also from context, general knowledge, 'only member of its class' etc.
- 2 Definite article being used both for definite determination and non-definite specificity
- 3 Article which is now used also for generics, a pure marker of nominality

Greenberg claims, by showing languages from all parts of the world, that the linguistic forms are at different stages. These stages are gradient, continuous and filled with intermediate stages, a much more fruitful hypothesis than the 'either-or' description by Krámský. It is an uncontroversial fact that the *definite article* has derived from the demonstrative determiner, and the *indefinite article* from the numeral 'one'. In several languages this derivation is in full bloom at present, and it is difficult to judge whether it has passed the 'borderline' yet or not. There are however quantitative differences between the occurrence of article vs. demonstrative, as shown by Cyr (ms.) among others.

But what about languages without an article, as in Type G. by Krámský and Stage 0 by Greenberg? Those are doubtless the most interesting ones for the purpose of this article. If one regards definiteness as a discourse-scope phenomenon one can discover still more structures. Below, I list some important means of expressing definiteness, where definiteness is seen as a wide discourse-scope phenomenon.

Definiteness can be expressed

| | |
|-----------------|---------------------------------------|
| Explicitly: | Article (independent words / affixes) |
| By other means: | Numeral for indefinite article |
| | Demonstrative for definite article |
| | Specific object marking |
| | Case marking |
| | Word order |
| | Theme particle |
| | Verbal aspect |

From the quite close similarity between articles and numerals/demonstratives the parallels become more and more vague. The last one, the aspect of a connecting verb, is a most *indirect* indicator of definiteness. This complex field will however not be treated here.

Definiteness in Greek, Swedish and Polish

In this chapter, the three languages of the study will be described, first in general terms and then related to the study. All three are Indo-European languages, spoken in Europe.

Forms of definiteness marking

Greek, forming a branch of its own of the Indo-European tree, is traditionally called a 'definiteness language', in so far that it has articles. The indefinite article shares its form with the numeral 'one', but it is

defined as a clear case of article. In speech, the difference between the indefinite article and the numeral is expressed by prosodic means. The near demonstrative independent pronoun shares form with the personal pronoun 'he'. As a demonstrative determiner, the pronoun + the article is used. Finally to be mentioned, the relation of possession has the article preposed and the possessive determiner postposed. All determiners inflect in two numbers, three genders and three to four cases.

Swedish, being a Germanic language, is also one of the traditional 'definiteness languages'. As for Greek, the indefinite article shares form with the numeral and is differentiated by prosodic means. The definite article has the form of a suffix, while the demonstrative determiners are independent and preposed. The possessive determiner is preposed. All determiners inflect in two numbers and two genders but no cases.

Polish, as a member of the Slavic branch, is traditionally considered as a 'non-definiteness language'. The numeral 'one' hardly serves as an indefinite article; in these cases the noun is bare. As for definiteness, the near demonstrative determiner does fill to some extent the function of a definite article in colloquial Polish, e.g. with the help of word order and prosody. All determiners inflect in two numbers, three genders and six to seven cases.

A summary of the definiteness marking is presented in (1).

(1) *Summary of definiteness marking in Greek, Swedish and Polish*

| | <i>Greek</i> | <i>Swedish</i> | <i>Polish</i> | |
|------------------|--------------|----------------|---------------|-------------|
| Indef. pron. | κάτι | något | coś | 'something' |
| Indef. det. | κάποιος N | någon N | jakieś N | 'some N' |
| Numerals | ένας N | en N | jeden N | 'one N' |
| | δυο N | två N | dwa N (ACC.) | 'two N:s' |
| | πέντε N | fem N | pięć N (GEN.) | 'five N:s' |
| Indef. art. | ένας N | en N | Ø N | 'a N' |
| Def. art. | ο N | N-en | Ø N (N ten) | 'the N' |
| Poss. det. | ο N μου | min N | mój N | 'my N' |
| | ο N του | Peters N | N Piotra | 'Peter's N' |
| | Πέτρου | | | |
| Pers. pron. | αυτός (drop) | han | on (drop) | 'he' |
| Rel. pron. | που | som | który | 'that' |
| Dem. det. (near) | αυτός ο N | den här N-en | ten N | 'this N' |
| (remote) | εκείνος ο N | den där N-en | tamten N | 'that N' |

Regarding the expression of possession, a small typological observation can be made. In Greek, all determiners are postposed. In Polish, pronouns are preposed and noun postposed. In Swedish, all determiners are preposed,

regardless of the number of 'owners'. This is illustrated in the scale below. See also English for comparison.

| | <i>Greek</i> | <i>Polish</i> | <i>English</i> | <i>Swedish</i> |
|-----------|--------------|------------------------------------|-----------------------------|----------------|
| PREPOSED | - | pronouns; nouns if sing. and human | pronouns | all |
| POSTPOSED | all | nouns | nouns if plur. or non-human | - |

Distribution of forms in the study

The corpus of the study consists of extracts from the Swedish children's book *Alla vi barn i Bullerbyn* by Astrid Lindgren and its translations into Greek and Polish. The language use is simple and not too distant from spoken discourse and syntax. From the three books, two extracts were chosen more or less at random.

All nouns and pronouns were registered and their forms were coded. Indefinite noun phrases were given the code ID with three degrees. ID1 denotes nouns with the indefinite article. ID2 contains noun phrases with numerals or other means to express bounded referents, e.g. accusative case marking. ID3 finally has those with indefinite pronouns or other means to denote nonreferential or unbounded referents, here genitive case marking.

| | | |
|-----|-----|---|
| ID1 | Gr. | Τα γατάκια νιαούριζαν μεσα σ' ένα καλάθι 'The kittens were miaowing in a basket' |
| ID2 | Sw. | Kristin har tre katter 'Kristin has three cats' |
| ID3 | Po. | Ugotowała nam też kawę 'She also made coffee for us' |

Bare nouns without any marking for definiteness were given a Ø-code.

| | | |
|---|-----|---|
| Ø | Po. | Kiedy jednak wyjrzałam przez okno... 'But when I looked through (the) window...' |
|---|-----|---|

For definite noun phrases, the code D was used, also in three degrees. D1 denotes nouns with definite article. D2 has possessive determiners. Finally, D3 contains demonstrative determiners, personal and relative pronouns. Also cases with pro-drop in Greek and Polish (compensated with verbal agreement) were given the code D3.

| | | |
|----|-----|--|
| D1 | Gr. | τα τρία γατάκια 'the three kittens' |
| D2 | Sw. | Norrgårdens gavel 'the gable of Norrgården' |

- D3 Po. (Ø) nie widziałam *go* nigdy przedtem
'I had never seen *it* before'

Those codes were arranged into a scale after the following principles. The no. 3s usually contain pronouns, ID3 denoting a non-existing or very vague referentiality and D3 a self-given one, even to the degree of pro-drop. For semantic reasons, they can be seen as extremes. The cognitive status of pronouns is self-evident and built into the very head. It is to be assumed that such categories are found in all languages.

From the extreme endpoints, the no. 2s draw nearer the center and towards an explicit marking of definiteness. They contain nouns with a determiner, usually numerals for ID2 and possessives for D2, but their cognitive status is less clear than that of the no. 3s.

The no. 1s are placed near the center. They consist of nouns with an article, ID1 with an indefinite one and D1 with a definite. Their cognitive status differs only according to the question of definiteness, i.e. if the referent is identifiable by the listener or not, and this status is expressed by the article, not by the noun.

In the very center is the bare noun, Ø, unmarked for definiteness. Here only the meaning of the word is important, not the property of identifiability.

(2) *Distribution of nouns with indefinite and definite determiners in the Greek, Swedish and Polish corpuses*

| | | Greek | Swedish | Polish |
|--------|--|-------|---------|--------|
| ID3: | Non-ref., Indef. pron., Gen. case | 7 | 6 | 8 |
| ID2: | Numerals, Acc. case | 10 | 12 | 11 |
| ID1: | Indef. art. | 14 | 19 | — |
| Ø: | Ø-art. | 7 | 49 | 95 |
| D1: | Def. art. | 90 | 28 | — |
| D2: | Poss. deter. | — | 14 | 14 |
| D3: | Pers. pron. (drop), Rel. pron., Dem. deter. | 82 | 82 | 82 |
| TOTAL | | 210 | 210 | 210 |
| ----- | | | | |
| Sum ID | | 31 | 37 | 19 |
| Ø: | | 7 | 49 | 95 |
| Sum D | | 172 | 124 | 96 |
| TOTAL | | 210 | 210 | 210 |

From the extreme outer values, with self-evident status and little explicit marking, the values towards the centre of the scale show a less self-evident split between indefinites and definites, which thus need an article, towards a bare noun, where the definiteness is zero. The results can be seen in (2).

Regarding the articles, the use of the indefinite article was moderate in the Greek and Swedish corpuses and non-existent in the Polish one. Bare nouns were rare in Greek, far more common in Swedish and extensive in Polish. The use of the definite article was extensive in Greek, moderate in Swedish and non-existent in Polish. Results like this are the reason for the label of Polish as a 'non-definiteness language'. But as will be seen in this paper, definiteness goes beyond the use of articles.

As for the rest, one can observe a much higher correspondence between the other determiners. Further, the results question the existence of any binary dichotomy of 'definiteness languages' vs. 'non-definiteness languages'. Greek and Swedish do have articles and should belong to the first group, but as can be seen, the distribution is highly gradient. A more suitable description would be that Greek has an extremely high occurrence of definite marking and Polish an extremely low one. Swedish is placed between the two.

Translation of forms

This chapter presents how the translations in the three languages correspond to the original. The various occurrences of determiners in an original text are expected to appear also in translations. It is to be assumed that definiteness in one language has usually received *some* marking of definiteness in the other ones.

The Swedish text is the source language, and both the Greek and the Polish texts are translated from this. It is a well known fact that the language use in translations differs slightly from that of native texts, due to 'translationese'. It is to be assumed that a similar text originally written in Greek and Polish respectively would differ slightly from the present translations. I am aware of this fact, and hopefully the findings give enough information despite this reservation (3).

The results show an extensive correspondence between the forms. One can observe that the endpoint forms are translated almost completely to corresponding forms (ID3 to ID3, D3 to D3). This is also valid for ID2. The closer we get to the center however, the less predictable are the translations. There are four striking deviations (*italics in the table*):

(3) Correspondence between the forms of definiteness in the Swedish original text and the Greek and Polish translations

| Forms of noun phrases in source language: | Forms of the same noun phrases in target languages: | | |
|--|--|--------------------------------|-----------------------|
| Swedish | | Greek | Polish |
| ID3: Non-ref., Indef. pron., Gen. case | 6 | ID3 6 | ID3 6 |
| ID2: Numerals, Acc. case | 12 | ID2 10 D1 1 D3 1 | ID2 11 Ø 1 |
| ID1: Indef. art. | 19 | ID1 14 ID3 1 Ø 1 D3 1 | Ø 17 D3 1 |
| Ø: Ø-art. | 49 | D1 42 Ø 6 D3 1 | Ø 47 ID3 1 D3 1 |
| D1: Def. art. | 28 | D1 27 D3 1 | Ø 26 D2 2 |
| D2: Poss. deter. | 14 | D1 14 Ø 2 | D2 12 |
| D3: Pers. pron. (drop), Rel. pron., Dem. deter. | 82 | D3 78 D1 4 | D3 80 Ø 2 |
| TOTAL | 210 | 210 | 210 |

1. Of the Swedish Øs, 42 are translated to D1 in Greek. This is easily explained, as 20 of the 42 items are proper names, 15 kinship terms and 2 generics, of which none have the definite article in Swedish but do in Greek. Naturally, proper names and kinship terms are overrepresented in this corpus, where relatives and playmates are prominent figures. A short glimpse at the use of the definite article in Greek indicates that it is used for many more subcategories of nouns than for instance in Swedish, namely for proper names, kinship terms, nouns with possessives, generics, etc..

2. All of the Swedish D2s are translated to D1 in Greek. This is also easily explained, as D2 denotes possessives, which all have the definite article in Greek. Moreover, this is a glide *inside* D, from one D category to another.

3-4. 17 ID1s and 26 D1s are translated to Ø in Polish. ID1 and D1 represent nouns with their whole definite property expressed only by their articles. Articles are certainly lacking in Polish, but as will be shown, the referential status can be compatible to those of other forms. A further analysis is made in the next section.

Referential status

This section will discuss the fact that the forms correspond to subgroups of varying referential prominence. Both Prince's scale, Givón's files and the Givenness hierarchy by Gundel et al. serve as encouraging starting points. Prince distributes entities into New-Inferrable-Evoked. Givón shows the separation of INDEFINITE into REFERENTIAL and NON-REFERENTIAL. But it has to be stressed – which Givón does – that not all DEFINITE are REFERENTIAL. Furthermore, the scale is much too coarse, as the stages corresponding to functions in languages are more than those three. The hierarchy by Gundel et al. is also a very useful point of departure in showing a scale of cognitive prominence related to the forms. But there are other instances of determination, such as possessives, that are not mentioned.

Low degree of referentiality

With these considerations as a background I will take a closer look at the various means of expressing definiteness, correlated to referential properties. The first major issue to be discussed is the property of *referentiality*. The lowest degree of referentiality is its very absence, with non-factive reference, i.e. in propositions within negated or hypothesised scope. The three languages use referential negation in these cases, Greek and Polish double negation and Swedish simple.

| | |
|-----|---|
| Gr. | <i>Καμία από τις μαμάδες δεν έφερε αντίρρηση...</i> |
| Sw. | <i>Ingen av mammorna brydde sej...</i> |
| Po. | <i>Żadna z mamus nie powiedziała nic...</i> 'None of the mummies had (not) any objection...' |

If the referent is in non-subject position, it has genitive case marking in Polish.

| | |
|-----|--|
| Gr. | <i>αλλιώς ... δε θα 'μενε χώρος για τη ίδια.</i> |
| Sw. | <i>annars skulle hon ... inte själv få rum att bo där.</i> |
| Po. | <i>inaczej ... dla niej samej nie starczyłoby w nim już miejsca.</i> GEN. CASE 'otherwise there would ... not be room for herself.' |

The same holds after semantically negative prepositions ('without', 'except' etc.) in Polish.

| | |
|-----|--|
| Gr. | <i>έπρεπε να τα δώσει όλα εκτός από ένα</i> |
| Sw. | <i>hon var tvungen att ge bort allihop utom en</i> |

- Po. muszona jest oddać wszystkie, *prócz jednego* GEN. CASE
'she had to give away all *except one*'

A higher degree of referentiality goes hand in hand with *boundedness*. Mass nouns with their unbounded characteristics are Ø-marked in all three languages. In Polish, in an object position, they take genitive case marking. See also Dahl & Karlsson 1975 for a study of a related language, Russian.

- Gr. και μας πρόσφερε και καφέ
Sw. Hon kokade kaffe också
Po. Ugotowała nam też kawę GEN. CASE
She also made *coffee* for us'

Plurals are usually considered more unbounded and less referential than singulars, and consequently, the forms of indefinite plurals and bare plurals coincide in the three languages, as well as in many other languages. An unbounded *number* of bounded plurals is usually expressed by indefinite determiners, with Polish using the genitive case.

- Gr. με πολλά πολλά μπουκέτα λουλουδία
Sw. med *en massa* små, små blombuketter
Po. z *masą* małych, małych *bukietów* kwiatów GEN. CASE
'with *many bunches* of flowers'

The use of numerals makes the plurals more bounded. In Polish, numbers from 5 and up require genitive case marking while numbers up to 5 require the accusative case. It is obvious that the lower the amount, the higher the referentiality.

- Po. dziewięć koron GEN. CASE
'*nine crowns* (Sw. currency)'
Po. trzy krzesła ACC. CASE
'*three chairs*'

Diffuse intermediate stages

The next issue is the question of *identifiability*, which constitutes the borderline between definiteness and indefiniteness and which needs deeper discussion. This is the group in which the three languages show the most widespread difference.

One subgroup covers entities which are *not* part of the hearer's discourse context, and therefore introduced during the ongoing conversation.

- Gr. βρισκόμουν σ' ένα δωμάτιο INDEF. ART.
Sw. Och då var vi i ett rum INDEF. ART.
Po. Znajdowaliśmy się w pokoju Ø
'We found ourselves in (a) *room*'

But there are numerous entities which are not necessarily referential but not completely imaginary either. The entities constitute a diffuse group extremely difficult to grasp.

A few words have to be said about generics. Generic reference is meant to cover all instances of a noun, e.g. 'man'. Its semantics are both definite and bounded, 'all people (of the earth)' and indefinite and unbounded 'any human being'. This supports the description of referentiality/definiteness as a hoop as in Givón 1978, 1984:407. Indisputably, this is not a question of a linear scale but of an ever-whirling wheel, where each end switches into its own opposite. Generic reference can be expressed in several ways in a language, but the most common one in Greek seems to be definite article + plural and in Swedish and Polish bare plural. The choice of plural indicates that referentiality is low and that this group might as well belong to the first parts of this scale, together with mass nouns and unbounded entities.

- Gr. 'Ολος ο κόσμος ... αγαπάει τα γατάκια DEF. ART. + SING. /
DEF. ART. + PLUR.
Sw. Alla människor ... tycker väl om kattungar QUANT. + PLUR. / PLUR.
Po. Wszyscy ludzie ... lubią chyba kocięta QUANT. + PLUR. / PLUR.
'All (the) *people* like (the) *kittens*'

Other entities are touched upon by Clark & Haviland 1977 in their discussion on 'shared knowledge' or by Hawkins 1978 on 'associative anaphoric'. They are 'stereotypically assumed' or 'inferrable', as labelled by Prince 1978, 1981. In Givón 1984, these entities hide both in his file 1 a) 'cultural entities' ('the president') and in file 2 b) 'relative deictic availability'. In file 2b), the entities under 'parts of whole' ('the kitchen') are most vague. In Gundel et al. 1993 they form an unclear middle stage between 'uniquely identifiable' and 'referential indefinite'. Maybe the best characteristic is the one given with the help of 'frame theory', expressed by Minsky 1975, to some extent also by Lakoff 1988 and others. Needless to say, this group is profoundly culturally dependent.

One subgroup consist of all entities which could be either definite or indefinite but with the same vague reference, e.g. 'the bank' or 'a bank'. They are traditionally labelled definite, but there are reasons to extend them to indefinites. Their referential status is drawn (inferred, associated...)

from the context (shared knowledge, culture...). However, the question of definiteness/indefiniteness is also a question of other factors. To illustrate the extremely subtle difference between the two, consider the scenarios below. Slightly depending on context, both a) and b) will do.

- a) See you later. I must go to *the bank* first.
 b) See you later. I must go to *a bank* first.
 a) I would like another size. – There is *the shop assistant*, ask her.
 b) I would like another size. – There is *a shop assistant*, ask her.

In Greek and Swedish the nouns within this group usually have an article – indefinite or definite – but it is not always easy to choose which one. In Polish they have Ø-form.

| | | |
|-----|--|-------------|
| Gr. | Ο μπαμπάς είχε κολλήσει την ταπεσαρία | DEF. ART. |
| Sw. | Pappa hade trollat dit tapeterna | DEF. ART. |
| Po. | Tatusz wyczarował prześliczne tapety | Ø |
| | 'Daddy had conjured up (the) wallpaper' | |
| Gr. | Ο μπαμπάς ... είχε φτιάξει ... ένα ... τραπέζι | INDEF. ART. |
| Sw. | Pappa hade ... trollat fram ... ett ... bord | INDEF. ART. |
| Po. | Tatusz ... wyczarowywał ... stolik | Ø |
| | 'Daddy had conjured up ... (a) ... table' | |

In the example above the only difference is obviously that wallpaper is an expected part of a room, while a table is not. Sometimes the choice differs between the two languages:

| | | |
|-----|---|-------------|
| Gr. | Μετά ξανανεβήκαμε τις σκάλες. | DEF. ART. |
| Sw. | ... om en stund gick vi upp för en trappa igen. | INDEF. ART. |
| Po. | a po chwili szliśmy znów w górę po schodach. | Ø |
| | '... after a while we climbed (the) (a flight of) stairs again' | |
| Gr. | Μετά μας έφτιαξε δίπλες στο ειδικό τηγάκι | DEF. ART. |
| Sw. | Och sen bakade Kristin rån åt oss i en tång | INDEF. ART. |
| Po. | Potem upiekła nam wafli w szczypcach | Ø |
| | 'And then K. baked wafers for us in (the) (a pair of) tongs' | |

A notice can be done about the examples above. 'Stairs' have plural form in Greek, Polish and English but not in Swedish. 'Tongs' have plural form in Polish and English but not in Greek and Swedish. Another indicator of their ambiguous referential status?

However, with more clues from the context the degree of referentiality increases, and the choice of indefinite article comes out of question. The hearer can more easily identify the referent, either by unique reference or inference. Examples of unique reference are natural phenomena and proper

names. Referentiality by inference is understood for kinship terms, parts-of-whole, body parts, (alienable) possession or when the referent has recently been mentioned and brought into the discourse.

| | | |
|-----|------------------------------------|-------------------|
| Gr. | ἡ Άννα | DEF. ART. |
| Sw. | Anna | Ø |
| Po. | Anna | Ø |
| | '(the) Anna' | |
| Gr. | Ἡ γιαγιά μου έβγαλε το μαντίλι | DEF. ART. |
| Sw. | tog mamma av mej handduken | DEF. ART. |
| Po. | mama zdjęła mi ręcznik | Ø |
| | 'mummy took off me (the) cloth' | |
| Gr. | μέσα στην κουζίνα | DEF. ART. |
| Sw. | i köket | DEF. ART. |
| Po. | w kuchni | Ø |
| | 'in (the) kitchen' | |
| Gr. | να ζεστάνουμε τα πόδια μας | DEF. ART. + POSS. |
| Sw. | värma våra fötter | POSS. |
| Po. | ogrzać nogi | Ø |
| | 'warm (the) (our) feet' | |
| Gr. | όλα τα βρεγμένα ρούχα μας | DEF. ART. + POSS. |
| Sw. | vi fick ta av oss våra våta kläder | POSS. |
| Po. | nasze przemoczone ubrania | POSS. |
| | '(the) (our) wet clothes' | |

As mentioned before, this is the group where the three languages really expose internal differences. As a general tendency, Greek uses definite article (*plus* possessive markers if applicable), Swedish has Ø-marking, definite article *or* possessives, and Polish has Ø-marking or possessives. Some triplet examples can be found in the corpus:

| | | |
|-----|---------------------------|-------------------|
| Gr. | στον κουμπαρά μου | DEF. ART. + POSS. |
| Sw. | i min sparbössa | POSS. |
| Po. | do skarbonki | Ø |
| | 'in (the) (my) money-box' | |

With this background it is time to return to table 3 and comment on the translation from Swedish to Polish. The nouns lost their articles in the translation process; 17 ID1 and 26 D1 were all translated into Ø.

The only difference between the use of the indefinite and the definite article is the property of identifiability. When looking at the referential status of the 17 ID1s, one discovers that only 6 are 'hard core' indefinites, i.e. previously unknown and introduced into the discourse context. The remaining 11 have ambiguous referential status. Some of them were

mentioned above as an illustration of this ('a table', 'stairs', 'tongs', 'a bank').

The same holds for the 26 D1s. Only 11 can be said to be easily identified through unique reference, recently mentioned, etc.. The rest have the same ambiguous status as the 11 IDs, where their reference is only indirectly inferred with a minimal boundary between 'expected' and 'non-expected' ('the wallpaper', 'the mats', 'the floor', 'the bank' etc.). One can easily see that the identifiability for these Swedish original cases is extremely vague, in spite of the article.

High degree of referentiality

Finally, when the entities have a self-evident reference, they are substituted with pronouns. This subgroup is called Evoked by Prince and In Focus by Gundel et al. The overwhelming part is held by personal pronouns but also relative pronouns and demonstrative determiners. As shown above, the use of pronouns is quite similar in the three languages with the exception that Greek and Polish allow pro-drop but not Swedish.

| | | |
|-----|---|-------------|
| Gr. | (Ø) ἐνώσα σα να (Ø) βγήκαμε | PRO-DROP |
| Sw. | kände jag att vi var ute | PERS. PRON. |
| Po. | (Ø) poczułam, że (Ø) jesteśmy na dworze '(I) felt that (we) were outdoors' | PRO-DROP |

Taken together, all nouns in the corpus show strikingly regular patterns in their marking of definiteness. The functions are degrees of referentiality, where factivity, boundedness and identifiability play important roles. The forms are articles, determiners, pronouns, case marking. With the considerations discussed above, the referential subcategories of the items in the corpus can be listed below (4).

Discussion

The referential categorisation above has given the required concepts for a general discussion of definiteness. The differences between the three languages of the study, Greek, Polish and Swedish, are numerous, but hopefully these results can illustrate their similarities.

The basis of the whole issue is *referentiality*, with some subproperties, among others *factivity*, *boundedness* and *identifiability*. The two extreme referential endpoints give a similar expression in all three languages. Of those, the first endpoint, non-factivity with lack of referentiality, is generally expressed with various kinds of negation. The other endpoint,

referentiality in focus, gives the personal pronouns. But it is the intermediate levels that are subject to typological variation.

(4) Referential subcategories in Greek, Swedish and Polish

| Referential status: | Expressed by: | | |
|---|---|---|--|
| | Greek | Swedish | Polish |
| 1. Non-factive reference | negation | negation | negation gen. case |
| <i>Unbounded reference</i> | | | |
| 2. - mass nouns | Ø | Ø | gen. case |
| 3. - unbounded amount | indef. pron. plural | indef. pron. plural | indef. pron. plural, gen. case |
| 4. Bounded amount | numerals | numerals | numerals gen. case ≥ 5 acc. case < 5 |
| 5. Reference not identifiable by the listener | | | |
| 6. Generic reference | indef. article definite article and plural | indef. article plural | Ø plural |
| <i>Reference identifiable by the listener</i> | | | |
| 7. - proper names | definite article | Ø | Ø |
| 8. - kinship terms | definite article | Ø | Ø |
| 9. - recently mentioned | definite article | definite article | Ø |
| 10. - parts-of-whole | definite article | definite article | Ø |
| 11. - body parts | definite article and possessives | definite article or possessives | Ø or possessives |
| 12. - alienable | definite article and possession | definite article or possessives | possessives possessives |
| 13. In focus | pers. pron. (drop), demonst. pron. relative pron. | pers. pron. demonst. pron. relative pron. | pers. pron. (drop) demonst. pron. relative pron. |

It seems that for Greek, the most important borderline is the one of *identifiability*, whether it is *possible* to identify the referent or not. If not possible, Greek uses indefinite markers, i.e. for entities that have obviously not yet been introduced into the discourse universe. But as soon as ever possible, the definite article is inserted, through several stages to the point where the nouns are substituted with pronouns. Greek avoids bare nouns.

For Swedish too, the crucial point is *identifiability*, but this language makes the choice at a later point than Greek. Swedish waits to insert the definite article until it is *more* necessary to identify the referent. If an indefinite article is not applicable, Swedish uses bare nouns much more than Greek and is also satisfied with mere possessive determiners. Both can be called 'identifiability languages', Greek to a higher degree and Swedish to a lesser.

For Polish, the crucial points are *factivity* and *boundedness*. Their means of expressing non-factivity/unboundedness, the genitive case marking, is kept to items less than 5 in number. Then, their nouns are Ø-marked until a considerable degree of referentiality, namely ownership, where possessive determiners are being used. Polish can be called a 'boundedness language'.

Further, it seems that between really unidentifiable and really identifiable nouns there is a large 'grey zone'. In Swedish and Greek, the presence of an article is obligatory, but the choice is ambiguous and often arbitrary. Only by indirect fine grained means, cultural knowledge and finger-tip feelings is it possible to make the right judgement. It happens that Greek and Swedish make a different choice: Greek chooses definite and Swedish indefinite article. The Polish nouns are here Ø-marked.

The sketched tables above hopefully gain in clarity, giving general lines and tendencies. They point to the important fact that a grammatical categorisation doesn't need to be binary or dichotomised. They illustrate that interlinguistic variations can group themselves into a larger pattern of universal regularity, and that typology, as well as other kinds of categorisation, doesn't need to contain a closed set of classes. The division into identifiability languages and boundedness languages perform an interplay of many factors.

Summary

After a short presentation of previous theoretical treatment of definiteness and some typological overviews, the main part of the paper deals with an analysis of Greek, Polish and Swedish. Definiteness, being a discourse-scope phenomenon, cannot be described only with attention to the forms but also with consideration to referential status, which in turn depends on many other factors.

It was first found that the split into 'definiteness languages' and 'non-definiteness languages' is highly questionable. The languages exhibit a gradient marking of definiteness. Especially between clearly indefinite and clearly definite nouns, there is a large number of entities with a very vague referential status and ambiguous marking. The languages also seem to pay attention to different factors of referentiality: identifiability in Greek (to a higher degree) and Swedish (to a lesser), factivity and boundedness in Polish. They can therefore be labelled identifiability language and boundedness language respectively, however without dichotomisation.

Yet there are still more questions than answers. An interesting field would be the role of generics, the role of pronouns, when they refer to 1st, 2nd or 3rd person, to human or impersonal entities, etc.. What are the connections to other linguistic expressions such as case marking, word order, verbal aspect? What is the relationship to other semantic functions such as theme/rheme, active/passive voice, background/foreground, etc.? These questions, which are also the subject of the author's current dissertation, need considerable exploration, a suitable subject for further research.

References

- Christopherson, P. 1939. *The articles: a study of their theory and use in English*. Copenhagen: Munksgaard.
- Clark, Herbert H. & Susan E. Haviland. 1977. 'Comprehension and the given-new contract'. In R. Freedle (ed.), *Discourse production and comprehension*, 1-40. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cyr, Danielle. ms. 'Crosslinguistic quantification: Definite article vs demonstratives'.
- Dahl, Östen & Fred Karlsson. 1975. *Verbal aspects and object marking. (Logical grammar reports 17)*. Göteborg: Dept. of Linguistics, University of Göteborg.
- Fraurud, Kari. 1994. ms. Paper presented at the Conference on anaphoric relations and (in)coherence, Antwerp, 1-3 December 1994.
- Givón, Talmy. 1978. 'Definiteness and referentiality'. In J.H. Greenberg (ed.), *Universals of human language. Vol. 4. Syntax*, 291-330. Stanford: Stanford University Press.
- Givón, Talmy. 1984. *Syntax. A functional-typological introduction I*. Amsterdam: John Benjamins.
- Greenberg, Joseph H. 1978. 'How does a language acquire gender markers?' In J. H. Greenberg (ed.), *Universals of human language. Vol. 3. Word structure*, 47-82. Stanford: Stanford University Press.
- Gundel, Jeanette, Nancy Hedberg & Ron Zacharski. 1993. 'Cognitive status and the form of referring expressions in discourse'. *Language* 69:2, 274-307.
- Haspelmath, Martin. 1994. 'Implicational universals in the distribution of indefinite pronouns'. *Sprachtypologie und Universalieforschung (STUF)* 47:3, 160-85.

- Hawkins, John. 1978. *Definiteness and indefiniteness*. London: Croom Helm.
- Heim, Irene. 1983. 'File change semantics and the familiarity theory of definiteness'. In R. Bäuerle, C. Schwarze & A. von Stechow (eds.), *Meaning, use and interpretation of language*, 164-89. Berlin: Walter de Gruyter.
- Karttunen, Lauri. 1976. 'Discourse referents'. In J. McCawley (ed.), *Syntax and Semantics 7*, 363-85. New York: Academic Press.
- Krámský, Jiří. 1972. *The article and the concept of definiteness in language*. The Hague: Mouton.
- Lakoff, George. 1988. 'Cognitive semantics'. In U. Eco, M. Santambrogio & P. Violi (eds.), *Meaning and mental representations*, 119-54. Bloomington: Indiana University Press.
- Lindgren, Astrid. 1948, 1995. *Alla vi barn i Bullerbyn*. Stockholm: Rabén & Sjögren.
- Lindgren, Astrid. 1961, 1981. *Τα έξι διαβολάκια του μικρού χωριού* (transl. by Ντίνα Κόμπα). Athens: Μετοπή Επε.
- Lindgren, Astrid. 1988. *Dzieci z Bullerbyn* (transl. by Irena Wyszomirska). Warsaw: Nasza Księgarnia.
- Lindvall, Ann. 1994. 'Aspect and species – a comparison between Polish and Swedish'. *Working Papers 42*, 103-15. Dept. of Linguistics, Lund University.
- Lindvall, Ann. forthc. 'Aspect and definiteness – a study of modern Greek'. In *Proceedings of 2nd international conference on Greek linguistics*. To be published by Benjamins.
- Minsky, Marvin. 1975. 'A framework for representing knowledge'. In P. H. Winston (ed.), *The psychology of computer vision*, 211-77. New York: McGraw-Hill.
- Prince, Ellen. 1978. 'On the function of existential presupposition in discourse'. In D. Farkas, W. Jacobsen & K. Todrys (eds.), *Papers from the Fourteenth Regional Meeting of the Chicago Linguistic Society*, 362-376. Department of Linguistics, University of Chicago.
- Prince, Ellen. 1981. 'Toward a taxonomy of given-new information'. In P. Cole (ed.), *Radical pragmatics*, 223-55. New York: Academic Press.

On noun-verb compounding in Swedish

Ingmarie Mellenius

Introduction

Within research on morphology, considerable effort over the last decades has gone into studying compounds, especially the more structural aspects of compounding: just as phrases have heads, words have been described as having heads, and the internal structure of a compound has been represented by the same type of tree diagrams with labelled nodes as are used to describe phrases (Selkirk 1982, Lieber 1983).

In languages like Swedish or English the overwhelming majority of actual compounds, and the majority of new coinages, are of the Noun-Noun type, in which there is no obvious, pre-determined way of analysing the semantic relationship between the modifying noun and the head noun.

In the case of synthetic compounds, i.e. compounds with a deverbal noun as head, on the other hand, the modifying element has been described as fulfilling an argument function, or as having a thematic role in relation to the verbal head, and the meaning of the compound is considered to be much more precise, and not as open to different interpretations. The differing semantic analyses for the two kinds of compounds are generally explained by stating that there is a crucial difference between the two word classes that function as the root morphemes of the heads in the respective constructions; a difference which can be expressed in different ways, e.g. by saying that whereas verbs are case-assigning items, nouns do not assign case.

The compound type that is the subject of this paper, Noun-Verb compounds, is considered to be very limited; it has even been claimed that constructions with a verb as head should be excluded from the universal set of word formation rules¹.

¹Wunderlich 1986, cited in Kiefer 1992b.