Grammaticalization in the nominal domain: the case of Polish cardinals[1]

Paweł Rutkowski
Warsaw University and Yale University

1 Introduction

In this paper, I propose a syntactic analysis of the diachronic development of Polish numeral expressions. It is observed that Q-numerals, such as *pięć* 'five', are functional elements derived from Old Polish nouns. Following a generative theory of language change put forward by Roberts and Roussou (1999), I interpret this shift from lexical (nominal) to functional status as a phenomenon driven by structural reduction, i.e. an example of grammaticalization. The syntax of numerals such as *pięćdziesiąt* 'fifty' or *pięćset* 'five hundred' is also analyzed as shaped by the force of simplification, however, in this case the source is not a single lexical element but a syntactic construction. I further argue that grammaticalization in the nominal domain is an on-going process: expressions with numerals such as *tysiąc* 'thousand' are undergoing structural simplification in present-day Polish.

2 Three types of cardinal numerals in Polish

Polish numerals are not a homogeneous class. The semantic set of cardinals can be divided into three distinct syntactic subclasses (cf. Rutkowski 2001, 2002a,

[1] I would like to acknowledge the financial support I received from the Polish-American Fulbright Commission during my stay at Yale University in 2005-2006. I am also grateful to Jadwiga Linde-Usiekniewicz, Ljiljana Progovac, Ian Roberts, Andrew Dombrowski, Paweł M. Nowak and Corey "Korzej" Yoquelet for comments on earlier versions of the analysis presented in this paper. Usual disclaimers apply.

Rutkowski and Szczegot 2001; see also Neidle 1988, Franks 1995, Giusti and Leko 1996 for similar classifications proposed for other Slavic languages):

- *A-numerals* (adjectival numerals) – the four lowest numerals (*jeden* 'one', *dwa* 'two', *trzy* 'three' and *cztery* 'four')

- *N-numerals* (nominal numerals) – very large numerals such as *tysiąc* 'thousand', *milion* 'million', *miliard* 'billion' etc.

- *Q-numerals* (numerals proper) – numerals such as *pięć* 'five', *piętnaście* 'fifteen', *pięćdziesiąt* 'fifty' or *pięćset* 'five hundred' (this is the biggest subclass).

These three subclasses differ in terms of case assignment. N-numerals resemble nouns because they always assign genitive to the quantified noun. Q-numerals require that the noun take genitive only when the larger nominal expression is in a structural case (nominative or accusative) position. In the context of inherent cases (genitive, dative, locative and instrumental)[2], Q-numerals agree in case with the noun. Finally, A-numerals agree with the quantified noun in all case contexts. These three patterns of morpho-syntactic behaviour are illustrated below: note that the verb *lubić* 'like' assigns accusative, whereas the verb *doradzać* 'advise' requires dative.

N-numerals:

(1a) *Cezary lubi milion osób.* [structural case context]
Cezary likes million-ACC people-GEN
'Cezary likes one million people.'

(1b) **Cezary lubi milion osoby.*
Cezary likes million-ACC people-ACC

(2a) *Cezary doradza milionowi osób.* [inherent case context]
Cezary advises million-DAT people-GEN
'Cezary advises one million people.'

(2b) **Cezary doradza milionowi osobom.*
Cezary advises million-DAT people-DAT

[2] For a more detailed discussion of the structural/inherent case distinction in Polish, see, e.g., Franks (1995), Przepiórkowski (1996), Rutkowski (2002a).

Q-numerals:

(3a) *Cezary lubi pięć osób.* [structural case context]
Cezary likes five-ACC people-GEN
'Cezary likes five people.'
(3b) **Cezary lubi pięć osoby.*
Cezary likes five-ACC people-ACC
(4a) *Cezary doradza pięciu osobom.* [inherent case context]
Cezary advises five-DAT people-DAT
'Cezary advises five people.'
(4b) **Cezary doradza pięciu osób.*
Cezary advises five-DAT people-GEN

A-numerals:

(5a) *Cezary lubi trzy osoby.* [structural case context]
Cezary likes three-ACC people-ACC
'Cezary likes three people.'
(5b) **Cezary lubi trzy osób.*
Cezary likes three-ACC people-GEN
(6a) *Cezary doradza trzem osobom.* [inherent case context]
Cezary advises three-DAT people-DAT
'Cezary advises three people.'
(6b) **Cezary doradza trzem osób.*
Cezary advises three-DAT people-GEN

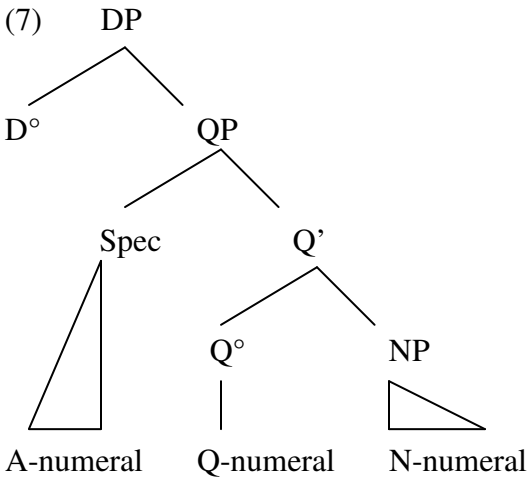
The following table summarizes this complicated pattern of case assignment:

Table 1: Genitive assignment in Polish numeral expressions

Genitive assignment	N-numerals	Q-numerals	A-numerals
in structural contexts	+	+	-
in inherent contexts	+	-	-

Rutkowski (2001, 2002a) attempts to account for this tripartite division by assuming that A-numerals are specifier-based modifiers (c.f., e.g., Giusti and Leko 1996, Veselovská 2001), and N-numerals have the syntactic status of nouns, whereas Q-numerals are functional elements, which are base-generated in a special projection (QP) in the region between DP and NP. The three possible syntactic locations of numerals are illustrated below:

(7)



This analysis can explain the complex pattern of case assignment in Q-type expressions. If functional (as opposed to lexical) elements are inserted into the syntax after inherent case assignment but before structural case assignment, their inability to assign case in the inherent case contexts is straightforward (the noun has already been assigned an inherent case value). Thus, Q-numerals, being functional, can only assign genitive in structural contexts (see Veselovská 2001, Rutkowski 2001, 2002a, for a more detailed analysis).

3 Q-numerals are grammaticalized nouns

The mixed pattern of case assignment in Q-type numeral expressions is a relatively recent innovation in Polish. In the 15th and 16th centuries, the equivalents of today's Q-numerals assigned genitive in both structural and inherent contexts (see, e.g., Klemensiewicz, Lehr-Spławiński and Urbańczyk 1964):

(8a) *siedm grzechow* [Old Polish - structural case context]
seven-NOM sins-GEN
'seven sins'

(8b) **siedm grzechy*
seven-NOM sins-NOM

(9a) *siedmią grzechow* [Old Polish - inherent case context]
seven-INSTR sins-GEN
'seven sins'

(9b) **siedmią grzechy*
seven-INSTR sins-INSTR

The above data show that Old Polish numerals such as *siedm* 'seven' behaved like regular nouns or N-numerals (*tysiąc* 'thousand' etc.). It is important to stress that, in Modern Polish, when expressions containing Q-numerals are sentential subjects, they do not agree with the verb – instead, the verb assumes a "neutral" form (third person neuter singular). As shown by Krasnowolski (1897), Szober (1923) and Schenker (1971), among others, there is both inflectional and syntactic evidence that Q-numerals in examples such as the following are accusative rather than nominative:

(10) Siedmiu rajtarów spało.
seven-ACC cavalrymen-GEN slept-3SG.NEUT
'Seven cavalrymen were sleeping.'

Some researchers have argued that it is the accusative status of subject Q-numerals that causes the lack of agreement with the verb – cf. Franks 1995, Przepiórkowski 2004, Rutkowski, 2000. Interestingly, Old Polish numerals such as *siedm* 'seven' were regular feminine nouns – they agreed with the verb and were unambiguously nominative, which is indicated by the inflectional form of the demonstrative in the following examples (cf. Klemensiewicz, Lehr-Spławiński and Urbańczyk 1964:401):

(11) Ona siedm panien szła. [Old Polish]
that-NOM.FEM seven-NOM maidens-GEN walked-3SING.FEM
'Those seven maidens were walking.'

(12) Tamte siedem panien szło. [Modern Polish]
those-ACC.FEM seven-ACC maidens-GEN walked.3SING.NEUT
'Those seven maidens were walking.'

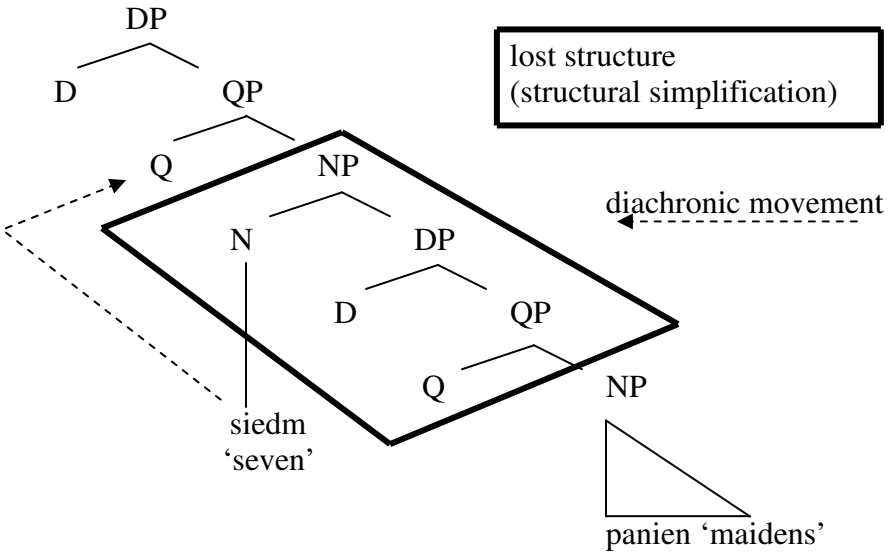
In this way, the syntactic behaviour of the Old Polish numeral *siedm* 'seven' was not different from feminine group nouns such as *grupa* 'group':

(13) Tamta grupa panien szła. [Modern Polish]
that-NOM.FEM group-NOM maidens-GEN walked-3SING.FEM
'That group of maidens walked.'

Therefore, there is no reason to assume that the Old Polish equivalents of today's Q-numerals were functional elements. The change from lexical to functional status has occurred between the 16th century and the present. This diachronic development patterns with what Roberts and Roussou (1999) consider grammaticalization, i.e. the reanalysis of lexical material as functional material. This model of grammaticalization assumes that the phenomenon in question

involves structural simplification – with a biphrasal expression becoming monophrasal. This is exactly what has happened in the historical evolution of Polish numerical structures. Old Polish numeral expressions consisted of two regular nouns; both of them projected full NPs and DPs. In Modern Polish, due to the N-to-Q (lexical-to-functional) shift of the numeral, these two extended projections have been reduced to only one (see Rutkowski 2002b, for a more detailed discussion):

(14) Diachronic loss of structure in Polish Q-type numeral expressions



As a result of this structural simplification, the case assigning properties of Q-numerals have changed: as mentioned in the previous section, Modern Polish cardinals such as *siedem* 'seven' are functional elements, which makes them unable to assign genitive in inherent case contexts.

It might be observed that, cross-linguistically, the phenomenon of grammaticalization is very often accompanied by processes of phonological erosion, adaptation and assimilation, as well as by the loss of independent morphosyntactic status of the grammaticalized elements (which often leads to morphological fusion) – cf., e.g., Croft (2000), Lehmann (1982). These processes were also present in the historical development of Polish Q-numerals. Some of those morphophonological changes started much earlier than syntactic grammaticalization. In Old Slavic, cardinalities such as '11' or '12' were expressed by means of complex syntactic constructions:

(15) jedinъ na desęte [Old Slavic]
one on ten
'eleven'

(16) dъva na desęte [Old Slavic]
two on ten
'twelve'

(17) tri na desęte [Old Slavic]
three on ten
'thirteen'

In Old Polish, these numerical expressions became simplified phonologically but not syntactically. As in Old Slavic, it was only the first element (the head of the whole construction) that declined, whereas the PP headed by the preposition *na* 'on' remained undeclinable in the eroded form *naście* (cf. Klemensiewicz 1974: 111). This is illustrated below (note that *z* 'with' is an instrumental assigner):

(18) siedm- na-ście wsi [Old Polish]
seven-NOM on ten villages-GEN
'seventeen villages'

(19) z siedmią- na-ście wsi [Old Polish]
with seven-INSTR on ten villages-GEN
'with seventeen villages'

If the head numeral was an A-numeral (the distinction between the four lowest numerals and the rest of the cardinal set was present as early as in Old Slavic), it agreed with the quantified noun – the noun not being assigned genitive:

(20) dwiema- na-ście wsiom [Old Polish]
two-DAT on ten villages-DAT
'twelve villages'

The syntactic structure of the above expressions should be represented in the following way (with the PP as a kind of adjunct):

(21) Num (PP) N

However, in Modern Polish, cardinalities such as '12' are expressed with simplex Q-numerals:

(22) siedemnaście wsi [structural context]
seventeen-ACC villages-GEN
'seventeen villages'

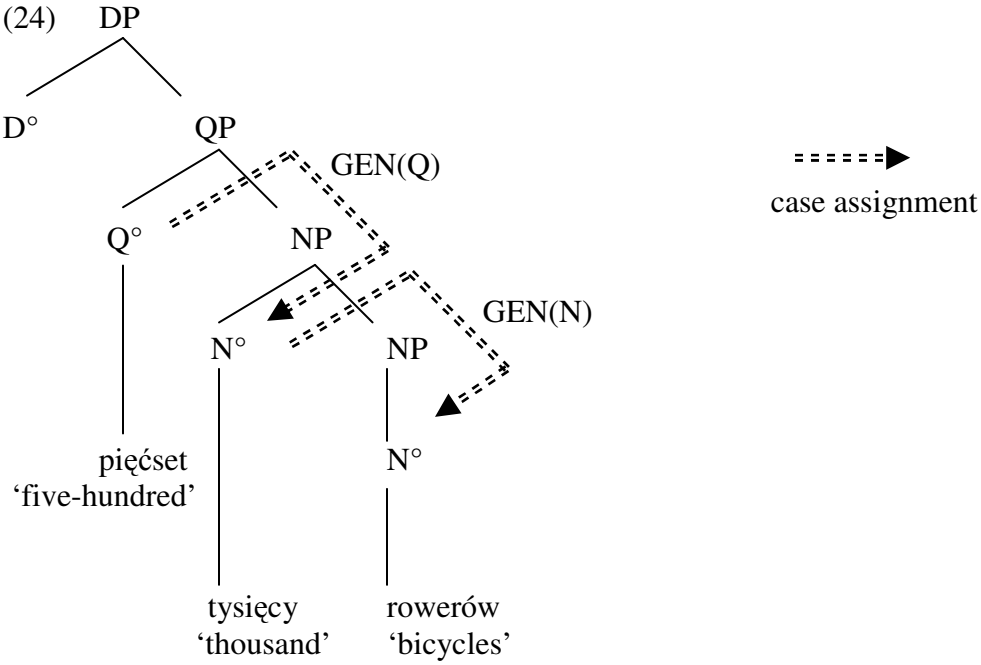
(23) z siedemnastoma wsiami [inherent context]
with seventeen-INSTR villages-INSTR
'with seventeen villages'

This means that the Num PP sequence has been syntactically reanalyzed: its elements are no longer independent, they have been fused. As with other Q-numerals, the case marking on the counted noun depends on the case context (structural or inherent).

4 Grammaticalization of N-numerals

The phenomenon of syntactic grammaticalization can also be traced in the historical development of complex numerical structures containing N-numerals. Rutkowski and Maliszewska (2006) analyse Modern Polish N-numerals as lexical heads projected within the same DP as the quantified noun:

(24)



The assumption that N-numerals do not project higher functional layers finds support in the fact that they cannot be pre-modified. I assume that adjectival

modifiers are hosted in functional phrases above the modified noun – if there is no functional material above N-numerals, no pre-modification can be possible:

(25a) niecałe pięćset rowerów
incomplete-ACC five-hundred-ACC bicycles-GEN
'less than five hundred bicycles'
(25b) niecałe pięćset tysięcy rowerów
incomplete-ACC five-hundred-ACC thousands-GEN bicycles-GEN
'less than five hundred thousand bicycles'
(25c) *pięćset niecałych tysięcy rowerów
five-hundred-ACC incomplete-GEN thousands-GEN bicycles-GEN

The Q-numeral and N-numeral seem to be inseparable – they belong to the same numerical expression. However, the N slot occupied by the N-numeral should not be analysed as a functional position (as shown in Section 2, N-numerals are genitive assigners in both structural and inherent case contexts). The nominal status of N-numerals can be best observed in structures with personal pronouns. Polish personal pronouns (as opposed to regular nouns) always precede Q-numerals:

(26a) sześćset Francuzek
six-hundred-ACC Frenchwomen-GEN
'six hundred Frenchwomen'
(26b) ich sześćset
they-GEN six-hundred-ACC
'six hundred of them'

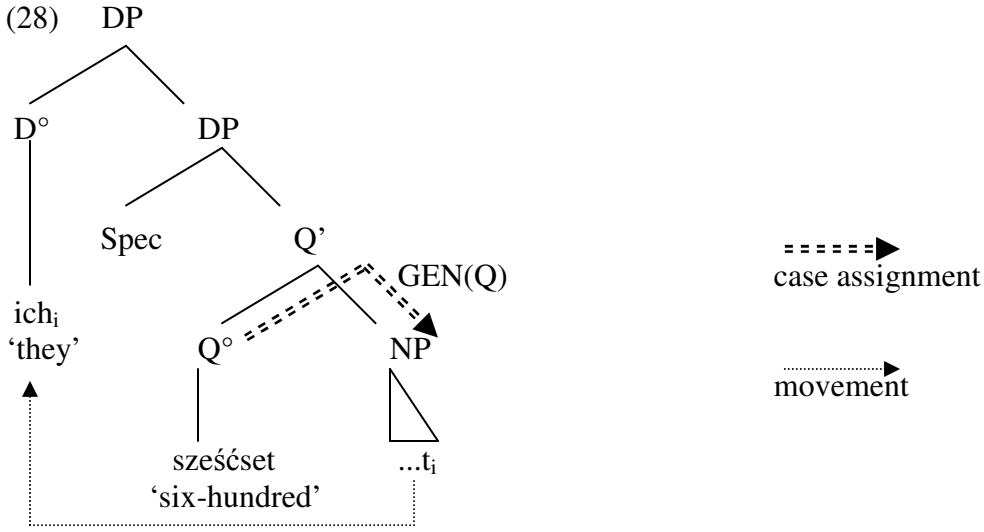
The above word order asymmetry finds a principled explanation if we assume that the pronoun is base-generated in N, assigned genitive, and then raised to D (for referential reasons) – see Rutkowski (2002c). However, in structures with N-numerals, the N-to-D movement of the personal pronoun is not possible:

(27a) sześćset tysięcy Francuzek
six-hundred-ACC thousands-GEN Frenchwomen-GEN
'six hundred thousand Frenchwomen'
(27b) *ich sześćset tysięcy
they-GEN six-hundred-ACC thousands-GEN

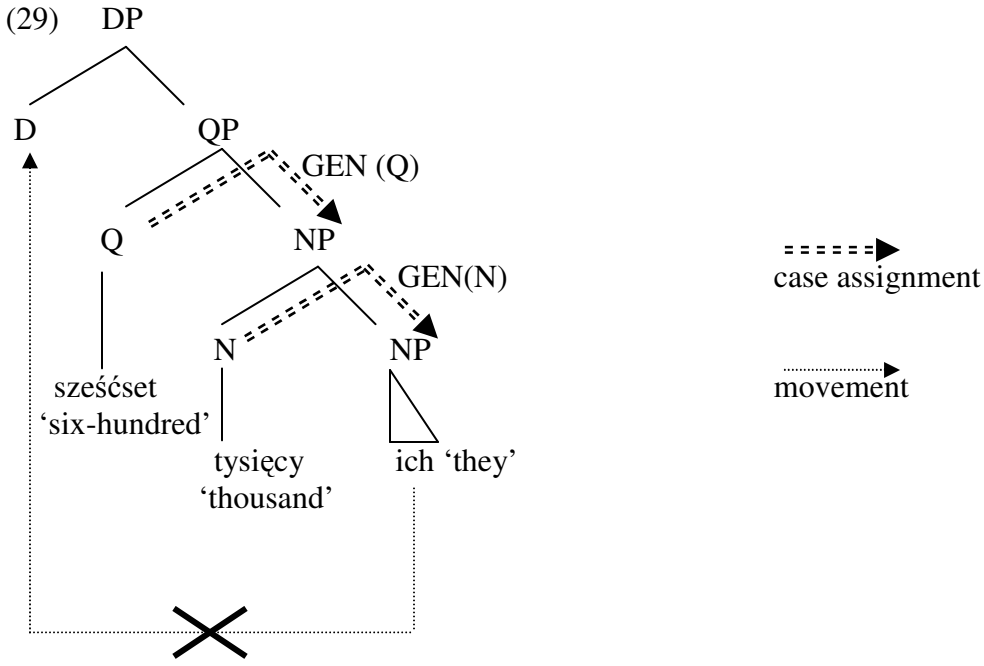
The ungrammaticality of (27b) would follow from the nominal status of the N-numeral: the personal pronoun does not move up because D can attract only the

closest N-type element (which is not a personal pronoun in this case). The derivations of examples (26b) and (27b) are shown in (28) and (29), respectively:

(28)



(29)



The unusual syntactic status of N-numerals (bare Ns projected between QP and NP) is reflected in their agreement properties. If the Q head is present, it obviously makes the verb assume the neutral form (see Section 3):

(30) Siedem tysięcy rajtarów spało.
seven-ACC thousands-GEN cavalrymen-GEN slept-3SG.NEUT
'Seven thousand cavalrymen were sleeping.'

However, even if the N-numeral is not preceded by a Q-numeral, the verb form is also third person neuter singular (although N-numerals such as *tysiąc* are morphologically masculine):

(31) Tysiąc rajtarów spało.
thousand-ACC cavalrymen-GEN slept-3SG.NEUT
'One thousand cavalrymen were sleeping.'

The above fact can be accounted for if, in structures such as (31), the N-numeral is analyzed as accusative rather than nominative (cf. Rutkowski 2000). If this analysis is on the right track, the phrasal status of N-numerals (i.e. their non-association with DPs) and their unusual case marking in the subject position (the same as the case marking of Q-numerals) can be explained as a result of a kind of grammaticalization process. I argue that Modern Polish N-numerals are gradually losing their nominal properties; they may be said to be half way between fully lexical and fully functional. They still assign genitive in all contexts but on the other hand, they are linked to the Q-type syntactic configuration. They resemble functional elements to a much greater extent than the Old Polish nouns from which they have developed. This process seems parallel to what happened to Old Polish complex numerals referring to numerosities such as '50' or '500'. They used to be syntactically analytic: they consisted of a cardinal and the noun meaning 'decade' or 'century' (cf. Szober 1923: 246-247). When the first element was an A-numeral, it agreed in case with the second element:

(32) cztery dziesięci rajtarów
four-NOM tens-NOM cavalrymen-GEN
'forty cavalrymen'

(33) trzy sta rajtarów
three-NOM hundreds-NOM cavalrymen-GEN
'three hundred cavalrymen'

However, when the first element was a higher cardinal (i.e. a syntactic noun), it assigned genitive to the element 'ten'/'hundred':

(34) pięć dziesiąt rajtarów
five-NOM tens-GEN cavalrymen-GEN
'fifty cavalrymen'

(35) pięć set rajtarów
five-NOM hundreds-GEN cavalrymen-GEN
'five hundred cavalrymen'

The above pattern matches the one found in Modern Polish structures with N-numerals. However, the Old Polish numerical complexes shown above have been reanalyzed in Modern Polish as simplex Q-numerals:

(36) pięćdziesiąt rajtarów
fifty-ACC cavalrymen-GEN
'fifty cavalrymen'
(37) pięćset rajtarów
five-hundred-ACC cavalrymen-GEN
'five hundred cavalrymen'

Numerical expressions with N-numerals such as *tysiąc* 'thousand' are unlikely to undergo a complete fusion because such a process of reanalysis would produce too many lexical entries (note that only eighteen new numerals evolved from the fused structures involving the nouns *dziesięć* 'ten' and *sto* 'hundred': '20'-'90' and '200'-'900'; in the case of structures with *tysiąc* 'thousand', far more combinations are possible). However, the process of syntactic simplification, as broadly understood, seems to be the same in both these cases. Therefore, I conclude that syntactic grammaticalization is a scalar phenomenon: in the development of Polish numerical expressions, structures with elements such as 'ten' or 'hundred' have been grammaticalized to a greater extent than structures with elements such as 'thousand'. Still, the latter are structurally simplified with respect to regular nominal expressions.

5 Conclusion

I hope to have shown that the diachronic development of the syntax of cardinal numerals between Old and Modern Polish is accounted for under the assumptions made by Roberts and Roussou (1999). I have employed their model of grammaticalisation to explain why Q-numerals are functional elements in Modern Polish, and how they evolved from Old Polish nouns. Structural simplification motivates primarily diachronic reduction of complex numeral expressions based on '10' and '100'. Finally, it has been observed that grammaticalization is in progress; N-type numeral expressions are undergoing syntactic simplification in Modern Polish.

References

Croft, William. 2000. *Explaining language change: an evolutionary approach*. Harlow: Pearson Education Limited.

Franks, Steven. 1995. *Parameters of Slavic morphosyntax*. Oxford, New York: Oxford University Press.

Giusti, Giuliana and Nedžad Leko. 1996. Definite and indefinite quantity expressions in Bosnian. In: *Determinatezza e indeterminatezza nelle lingue slave. Atti del convegno svoltosi a Firenze, 26-28 ottobre 1995*, ed. Rosanna Benacchio, Francesca Fici and Lucyna Gebert, 127-145. Padova: Unipress.

Klemensiewicz, Zenon. 1974. *Historia języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.

Klemensiewicz, Zenon, Tadeusz Lehr-Spławiński and Stanisław Urbańczyk. 1964. *Gramatyka historyczna języka polskiego*, 2nd edition. Warszawa: Państwowe Wydawnictwo Naukowe.

Krasnowolski, Antoni. 1897. *Systematyczna składnia języka polskiego*. Warszawa: Drukarnia Estetyczna K. Sierpińskiego.

Lehmann, Christian. 1982. *Thoughts on grammaticalization: a programmatic sketch*, vol. I. Köln: Institut für Sprachwissenschaft.

Neidle, Carol. 1988. *The role of case in Russian syntax*. Dordrecht, Boston: Kluwer Academic Publishers.

Przepiórkowski, Adam. 1996. Case assignment in Polish: towards an HPSG analysis. In *Studies in HPSG* (*Edinburgh Working Papers in Cognitive Science,* vol. 12), ed. Claire Grover and Enric Vallduví, 191-228. Edinburgh: University of Edinburgh.

Przepiórkowski, Adam. 2004. O wartości przypadka podmiotów liczebnikowych. *Biuletyn Polskiego Towarzystwa Językoznawczego/Bulletin de la Société Polonaise de Linguistique* 60:133-143.

Roberts, Ian and Anna Roussou. 1999. A formal approach to "grammaticalization". *Linguistics* 37:1011-1041.

Rutkowski, Paweł. 2000. Składnia polskich grup liczebnikowych: próba opisu formalnego. *Poradnik Językowy* 8:10-28.

Rutkowski, Paweł. 2001. Numeral Phrases in Polish and Estonian. In *Proceedings of the 18th Scandinavian Conference of Linguistics*, vol. 2, ed. Arthur Holmer, Jan-Olof Svantesson and Åke Viberg, 181-190. Lund: Lund University Press.

Rutkowski, Paweł. 2002a. The syntax of Quantifier Phrases and the inherent vs. structural case distinction. *Linguistic Research* 7.1:43-74.

Rutkowski, Paweł. 2002b. Numerals as grammaticalised nouns: a generative approach. *Interlingüística* 13.III:317-328.

Rutkowski, Paweł. 2002c. Noun/pronoun asymmetries: evidence in support of the DP hypothesis in Polish. *Jezikoslovlje* 3.1-2:159-170.

Rutkowski, Paweł and Hanna Maliszewska. 2006. On Prepositional Phrases inside numeral expressions in Polish. *Lingua* (to appear).

Rutkowski, Paweł and Kamil Szczegot. 2001. On the syntax of functional elements: numerals, pronouns and expressions indicating approximation. In *Generative linguistics in Poland: syntax and morphosyntax*, ed. Adam Przepiórkowski and Piotr Bański, 187-196. Warszawa: Instytut Podstaw Informatyki Polskiej Akademii Nauk.

Schenker, Alexander M. 1971. Some remarks on Polish quantifiers. *The Slavic and East European Journal* XV.1:54-60.

Szober, Stanisław. 1923. *Gramatyka języka polskiego*, 2nd edition. Lwów, Warszawa: Książnica Polska, Tow. Nauczycieli Szkół Wyższych.

Veselovská, Ludmila. 2001. Agreement patterns of Czech group nouns and quantifiers. In *Semi-lexical categories: the function of content words and the content of function words*, ed. Norbert Corver and Henk van Riemsdijk, 273-320. Berlin, New York: Mouton de Gruyter.

Warsaw University
Department of General Linguistics and Baltic Studies
Katedra Językoznawstwa Ogólnego i Bałtystyki
ul. Krakowskie Przedmieście 26/28
00-927 Warszawa
Poland

p.rutkowski@uw.edu.pl