1. The lexical-functional divide and semi-lexicality

Wikipedia: “The term **lexical category** applies only to those parts of speech and their phrasal counterparts that form open classes and have full semantic content. The parts of speech that form closed classes and have mainly just functional content are called **functional categories**”

Corver & van Riemsdijk (2001), Abney (1987), lexical vs. functional:

<table>
<thead>
<tr>
<th>Lexical:</th>
<th>Functional:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open class</td>
<td>Closed class</td>
</tr>
<tr>
<td>Phonologically and morphologically independent</td>
<td>Phonologically and morphologically dependent</td>
</tr>
<tr>
<td>Take different kinds of categories as their complement</td>
<td>Restricted to a single category as their complement</td>
</tr>
<tr>
<td>Assign theta roles</td>
<td>Do not assign theta roles</td>
</tr>
<tr>
<td>Separable from complement</td>
<td>Inseparable from complement</td>
</tr>
</tbody>
</table>

Example: verb (lexical) vs. plural morphemes (functional)
- Verbs are open class, plural morpheme part of a closed class [sg, pl]
- Verbs are phonologically & morphologically independent, while plural morpheme requires a noun
- Verbs take different types of complements (CP, DP, PP, etc.), while plural morphology only combines with nouns
- Verbs assign theta roles; plural morphology does not
- Verbs can be separated from their complements (what did he eat __?), while plural morphology cannot (*book, he bought s*)

But, the distinction is not perfect!

Example: English numerals (*one, five, eighty, thousand*)
- Numerals are usually considered closed class, but notice the ability to form new numerals (= lexical property):
  - Fictitious quantities: jillion, bajillion, zillion, gazillion
  - Very high quantities: googol ($10^{100}$), googolplex ($10^{10^{100}}$ ‘ten to the ten to the hundred’), novemdecillion ($10^{60}$), vigintillion ($10^{63}$)
- They are phonologically & morphologically independent (Lexical)
- They take one type of complement, namely noun (*three eat*) (Functional)
- They do not assign theta roles (Functional)
- They cannot be separated from their complements (*What did he buy three__?*) (Functional)

English numerals are lexical by two counts and functional by three!

Semi-lexicality: “Certain lexical items display ambiguous behavior: they share properties with lexical categories and at the same time they display functional characteristics.” (Corver & van Riemsdijk 2001: 3)

Roadmap:
- Introducing and formalizing the notions of lexical, functional, and semi-lexical
- Case study #1: Polish numeral 1000,
- Case study #2: English Q-nouns *lot, ton, bunch*
- Case study #3: Dutch kind-expression *soort*

2. What is semi-lexicality?

2.1 The lexical-functional landscape theoretically

Early theorizing: Structure building revolves around lexical categories, with functional elements falling in as satellites of the lexical category (PS rules), e.g.:

```
NP \rightarrow Det N
```

```
Det the
N girl
```
Later theorizing: Functional categories head their own phrases (X-bar theory), e.g.:

\[
\begin{array}{c}
\text{DP} \\
\text{D} \quad \text{NP}
\end{array}
\]

This has developed into more complex syntactic structures, with each functional category heading its own functional projection, e.g.:

\[
\begin{array}{c}
\text{DP} \\
\text{D} \quad \text{QP} \quad \text{NP}
\end{array}
\]

...which collectively have been referred to as an “extended projection” (e.g. Grimshaw 1991, or van Riemsdijk 1990 “categorial feature magnetism”) (see discussion in Corver (2013)).

Towards roots and features:

- Modern syntactic theory is in an “era of features”, where functional categories are realized by functional features (e.g. T = [past, pres], # = [sg, pl], etc.) and functional features drive syntactic computation (Merge, Agree, EPP).
  - E.g. as in Distributed Morphology (Halle & Marantz 1993, Marantz 1997), Cartography (Cinque 1999, 2010), Nanosyntax (Starke 2009), Minimalism (e.g. Adger 2003).
- These functional projections dominate a root (which in some approaches needs a category-defining functional head, n, v, or a), forming an extended projection.
- N, V, or A of early theories is thus a root, plus the set of functional projections needed to make it into a noun, verb, or adjective.
- Modern syntax consists of roots and functional categories / features.

2.2 Operationalizing semi-lexicality

Lexical: At the core of every noun, verb, or adjective is a root. I propose that what makes something lexical is the presence of a root.

Functional: Syntactic features encode grammatical notions like tense, number, animacy, and drive syntactic computation via the EPP, agreement probes, etc. I propose that what makes something functional is the presence of a feature.

Semi-lexicality Hypothesis: Semi-lexicality is what occurs when a root is lexically specified for a syntactic feature.

- Roots not specified for any features are predicted to be flexible, permitted wherever roots are permitted (e.g. brightA lights vs. his brightNs to chair a discussion, the chairs).
- Roots specified for features, however, will constrain how the syntactic structure embedding the root is constructed.

Toy example #1: Gender (γ) on nouns

Many languages divide nouns according to gender, e.g. as in Polish:

- Masculine: chłopiec ‘boy’, pies ‘dog’, stół ‘table’
- Feminine: dziewczyna ‘girl’, krowa ‘cow’, koszula ‘shirt’
- Neuter: dziecko ‘child’, bydło ‘beast’, krzesło ‘chair’

In most cases (human-denoting nouns aside), the gender specification represents arbitrary morphosyntactic information.

Under the approach to semi-lexicality here, that information is encoded in the lexical entry of the root, as a specified gender feature:

\[
\begin{array}{c}
\gamma P \\
\gamma \sqrt \text{P} \\
\text{M/F/N} \sqrt \text{noun}
\end{array}
\]

When a \( \gamma P \) is constructed above a root, the gender specification of the root will tell the syntax what value to assign that \( \gamma \) node. Note that in a gendered language like Polish, the specification of gender will only have the effect of valuing \( \gamma \).
Toy example #2: Pluralia tantum nouns

Many languages include a small subset of nouns which are necessarily specified for plurality, called pluralia tantum (see discussion in Acquaviva 2008).

(9) pants, scissors

Syntactically, these behave like plural nouns, appearing only in contexts where plurality is permitted. In English for example, they do not occur in mass or singular count contexts:

(10) a. *Much pants / scissors Mass context
    b. *a pants / scissors Singular count context

In addition, it is only when functioning as a lexical noun that these elements need to be plural – as verbs (e.g. to scissor), or in compounds (e.g. pant leg), the obligatory plurality does not apply.

Under the semi-lexicality hypothesis, these nouns are specified for plurality, requiring a plural #P when embedded in a nominal structure.

(11) Pants: [ √pants, PL #], scissors: [ √scissors, PL # ]

Unlike γ, the specification of # requires the projection of structure which otherwise is not obligatory, e.g. nouns can often also be mass (no #P), or singular (singular #P). The specification of # excludes such configurations.

3. Polish numeral 1000

Numerals in Polish are notorious for their semi-adjectival/nominal behavior. Much of the literature has focused on the 5+ numerals (5-10, 100), which in Polish and many other Slavic languages vary between a case-agreeing and case-assigning form:

(13) a. Pięć ptaków Five.NOM/ACC birds.GEN
    ‘Five birds’
  b. Z pięćoma ptakami WithINST five.INST birds.INST
    ‘with five birds’

The numeral 1000 functions as a base in Polish, and has drawn much less attention due to the fact that it lacks the idiosyncratic case properties of the 5+ numerals:

(14) a. Tysiąc ptaków Thousand.NOM/ACC birds.GEN
    ‘Thousand birds’
  b. Z tysiącem ptaków WithINST thousand.INST birds.GEN
    ‘With a thousand birds’

Instead, where addressed, in Polish and other Slavic languages, it and numerals higher have been termed “nouns”, “noun-like” (e.g. Corbett 1978, Baby 1987, Franks 1994, Rappaport 2003, Miechowicz-Mathiasen & Dzubala-Szrejbrowska 2012).

Despite this, it has non-nominal idiosyncrasies, which make it only a “semi-noun.” I show this below, drawing on the following data sources:
- Swan (2002), a comprehensive grammar of Polish
- Polish National Corpus (Przepiórkowski et al. 2011), examples marked with “NKJP” (Narodowy Korpus Języka Polskiego)
- Native speaker judgments / consultation

3.1 Properties of Polish 1000

The numeral looks nominal with regards to its paradigm (15) and case assignment properties (14), repeated as (17), but it shows idiosyncrasies in agreement ((18)-(24)).

PARADIGM: Polish 1000 inflects identically to a masculine inanimate noun (comparison word miesiąc ‘month’, chosen for its phonological similarity); it inflects for both singular and plural:

(15) a. Dziesięć ptaków 10 N.MEN/PL birds.PL
    ‘Ten birds’
  b. Z dziesięcioma ptakami WithINST 10.INST birds.INST
    ‘With ten birds’
(15) Paradigm of 1000 tysiąc, compared to miesiąc ‘month’

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>tysiąc</td>
<td>miesiąc</td>
</tr>
<tr>
<td>ACC</td>
<td>tysiąc-a</td>
<td>miesiąc-a</td>
</tr>
<tr>
<td>GEN</td>
<td>tysiąc-ą</td>
<td>miesiąc-ą</td>
</tr>
<tr>
<td>DAT</td>
<td>tysiąc-owi</td>
<td>miesiąc-owi</td>
</tr>
<tr>
<td>LOC</td>
<td>tysiąc-u</td>
<td>miesiąc-u</td>
</tr>
<tr>
<td>INST</td>
<td>tysiąc-em</td>
<td>miesiąc-em</td>
</tr>
</tbody>
</table>

Plurality: The plural form is used in complex numerals, or to indicate abundance.

(16) a. Tysiąc ptaków
   Thousands.NOM/ACC birds.GEN
   ‘Thousands of birds’

   b. Dwa tysiące ptaków
   Two.M/NOM/ACC thousand.NOM/ACC birds.GEN
   ‘Two thousand birds’

Case assignment: 1000 behaves similar to an argument-taking noun klucz ‘key’:

(17) a. Tysiąc ptaków / Klucz ptaków
   Thousand.NOM birds.GEN / Key.NOM birds.GEN
   ‘A thousand birds / A flock of birds (flying in a V)’

   b. Z tysiącem ptaków / kluczem ptaków
   With.NOM thousand.INST birds.GEN / Key.INST birds.GEN
   ‘With a thousand birds / a flock of birds (flying in a V)’

Agreement: The numeral also seems nominal when it comes to agreement. Agreement can target the numeral, surfacing as masculine (inanimate) singular.

(18) a. …i dobry tysiąc kamieni przemknął nam nad karkami.
   and good.M.SG thousand.NOM rocks.GEN flew.M.SG us over necks
   ‘and a good thousand rocks flew over our heads (lit. necks).’ (NKJP)

   b. Do Albanii dotarł pierwszy tysiąc żołnierzy
   To Albania reached.M.SG first.M.SG thousand.NOM soldiers.GEN
   sił międzynarodowych.
   force.GEN international.GEN
   ‘The first thousand soldiers from the international forces reached
   Albania.’ (NKJP)

c. Ponieważ ten tysiąc zawierał prawie całą
   Because DEM.M.SG thousand.NOM contained.M.SG almost all.ACC
   śmietankę polskiego towarzystwa...
   cream.ACC Polish.GEN society.GEN
   ‘Because this thousand contained almost all the cream of Polish society’
   (NKJP)

At the same time, the numeral is known to trigger default verbal agreement, which is realized with neuter singular features (Dziwirski 1990):

(19) a. Przez ciebie tysiąc kijów spadło na mój grzbiet!
   By you thousand sticks.GEN fell.N.SG on my back
   ‘Because of you, a thousand sticks fell on my back!’ (NKJP)

   b. Wokół niego tysiąc gwiazd migotało na niebie …
   Around him thousand stars.GEN flickered.N.SG on heavens
   ‘Around him, a thousand stars flickered in the heavens…’ (NKJP)

   • That this is default agreement can be seen by its inability to trigger plural verbal
     agreement when coordinated with another high numeral expression (as
     compared to coordinated neuter singular nouns):

(20) a. Tysiąc ptaków i milion motyli zginęło.
   Thousand birds.GEN and million butterflies.GEN perished.N.SG
   ‘A thousand birds and a million butterflies perished.’

   b. Krzesło i biuroko rozbiły się.
   Chair.N.SG and desk.N.SG broke.NV.PL SIE
   ‘A chair and a desk broke.’

   • …in combination with the fact that subject-less weather verb constructions
     involve neuter singular verbal agreement (=the default):

(21) Padalo.
    Rained.N.SG
    ‘It rained.’
Polish 1000 can also trigger default agreement on pre-modifiers (in combination with default verbal agreement):

(22) a. Te tysiąc złotych przeznaczone była na DEM.NV.PL thousand.NOM gold.GEN designed.NV.PL was.N.PL for jakiś cel i już zostało wydane. some purpose and already was.N.PL spent.NV.PL

‘These thousand gold(Polish currency) were designed for some purpose and have already been spent.’ (NKJP)
b. Niecałe tysiąc Koreańczyków wyemigrowało do Polski. Incomplete.NV.PL thousand Koreans.V.GEN emigrated.N.SG to Poland

‘Almost one thousand Koreans emigrated to Poland.’ (adapted from Rutkowski 2006)

• That this is default agreement is visible in (22)b: the form niecałe is non-virile plural, which requires a non-virile (= not masculine human, i.e. not ‘virile’), plural target. Tysiąc is non-virile (M.INAM), but singular, and Koreańczyków is plural, but virile; the features non-virile plural must represent a default.

Pre-modifiers can also agree with the quantified noun (in combination with default verbal agreement):

(23) a. Kolejnych tysiąc osób przyszło osobiście albo zadzwoniło. Another.GEN thousand people.GEN came.N.SG in-person or called.N.SG

‘Another thousand people came in person or called.’ (NKJP)
b. Że tych tysiąc górników od razu poszło się… that DEM.GEN thousand miners.GEN at once went.N.SG SIE ‘…that a thousand miners went at once (to…)' (NKJP)

Crucially, patterns of successful and default agreement do not seem to mix, e.g. M.SG verbal agreement correlates with M.SG pre-modifier agreement, and default verbal agreement correlates with default pre-modifier agreement or pre-modifier agreement with the noun:

(24) a. %Dodatkowy tysiąc osób dal-o-by się Additional.M.SG thousand people.PL allow-N.SG(DEF)-COND SIE upchnąć w przewyżshych pensjonatych. push.INF in private.LOC pensions.LOC ‘An additional one thousand people would allow themselves to be pushed into private pensions.’ (NKJP) (pattern rejected by n=4, of 5)

b. *Dodatkowe tysiąc osób dal-by się Additional.NV.PL(DEF) thousand people.PL allow.M.SG-COND SIE upchnąć w przewyżshych pensjonatych. push.INF in private.LOC pensions.LOC (pattern rejected by n=5, of 5)
c. *Dodatkowych tysiąc osób dal-by się Additional.NV.PL(DEF) thousand people.PL allow.M.SG-COND SIE upchnąć w przewyżshych pensjonatych. push.INF in private.LOC pensions.LOC (pattern rejected by n=1, of 1)

When plural, both default and successful verbal agreement is found. Note that identifying default pre-modifier agreement is no longer possible, as tysiące is both plural and non-virile, the default features DP-internally.

(25) a. Tysiące oczu gapia się Thousands eyes.GEN staring.PL SIE ‘Thousands of eyes were staring.’ (NKJP)
b. Tysiące ptąń przychodziło mi do głowy. Thousands questions.GEN came.N.SG to me to head ‘Thousands of questions came to my head.’ (NKJP)
c. …o tych tysiace uczniów, wychowanków… about DEM.PL.GEN thousands students.GEN pupils.GEN ‘…about those thousands of students, pupils…” (NKJP)

Putting this together, there are two patterns of agreement, Pattern 1 where agreement targets the numeral (M.SG or NV.PL), and Pattern 2, where verbal agreement defaults and pre-modifier agreement defaults or targets the noun.2


‘A whole thousand birds slept.’
b. Te tysiące ptaków spały. DEM.NV.PL thousand birds.GEN slept.NV.PL ‘Those thousands of birds slept.’

(27) a. Cale /calych tysiąc ptaków spało. Pattern 2 Whole.NV.PL /GEN thousand birds.GEN slept.N.SG DEFAULT

‘A whole thousand birds slept.’

1 A few more words are necessary on Pattern 1. Speakers usually require an agreeing pre-modifier to be present for verbal agreement with singular 1000 to be successful, e.g. (26a) without the adjective would be judged ungrammatical by most speakers. Furthermore, which pre-modifiers are allowed to agree with the numeral differs between speakers. Together, this means that Pattern 1 is much more restricted than Pattern 2, and where it is permitted will differ from speaker to speaker.
3.2 The semi-lexicality of Polish 1000

Summary:
- Numeral 1000 looks like a lexical noun, in terms of its paradigm and case assignment properties.
- But, numeral 1000 shows a mixed agreement paradigm, either controlling agreement like a lexical noun, or not.
- Conclusion: Numeral 1000 is a semi-lexical noun, or “semi-noun.”

3.2.1 Internal syntax of Polish 1000

Hypothesis: The two agreement paradigms correspond to two different syntactic instantiations of numeral 1000: 1000 as a lexical noun, or as a semi-lexical noun.

Further evidence: Polish has a distributive marker po, which triggers locative case on lexical nouns (see Przepiórkowski 2010, 2013 and Przepiókowski and Patujek 2013 for discussion and analysis), but behaves inert for case assignment with numerals (e.g. the numeral surfaces in the case of the case context).

(28) a. Dalam im po jabłku / *jabłko  
   Gave.1SG.F them DIST apple.LOC/*apple.ACC  
   ‘I gave them an apple each.’

   b. Dalam im po pięć jabłek / *pięciu jabłkach. Numeral  
   Gave.1SG.F them.DAT po five.ACC apples.GEN / five.LOC apples.LOC  
   ‘I gave them five apples each.’

Numeral 1000 shows optionality, patterning with either:

(29) Dalam im po tysiąc / *tysiącu jablek.  
   Gave.1SG.F them.DIST thousand.ACC / thousand.LOC apples.GEN  
   ‘I gave them a thousand apples each.’

This further supports the hypothesis that the numeral can have two syntactic instantiations.

Pattern 1 (full agreement):

(30) a. Calem / *calych tysiąc ptaków spalo.  
   Pattern 1  
   Whole.M.SG thousand birds.GEN slept.M.SG  
   ‘A whole thousand birds slept.’

   b. Te / *tuch tysiące ptaków spalo.  
   DEM.NV.PL / *tuch thousand birds.GEN slept.NV.PL  
   ‘Those thousands of birds slept.’

The numeral has phi-features, both # (SG or PL) and γ (masculine, inanimate). These act as an intervener for agreement, thereby leading to the numeral controlling both the modifier and verb.

Assuming both # and γ project in the syntax, the internal syntax of 1000 in Pattern 1 would have the following form:

<table>
<thead>
<tr>
<th>TP</th>
<th>DP</th>
<th>ADJ</th>
<th>…</th>
<th>NGEN</th>
<th>[γ, T</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>[γ, #, #GEN [γ, –y, –s]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pattern 2 (default agreement):

(33) a. Calem / calych tysiąc ptaków spalo.  
   Pattern 2  
   Whole.NV.PL / .GEN thousand birds.GEN slept.N.SG  
   ‘A whole thousand birds slept.’

   b. Te / *tuch tysiące ptaków spalo.  
   DEM.NV.PL / .GEN thousand birds.GEN slept.N.SG  
   ‘Those thousands of birds slept.’

The numeral is phi-feature deficient. It carries number (SG or PL), but no γ. Because the noun is also genitive, there is no target for successful verbal agreement; the pre-modifier can target the noun (genitive) or numeral (default).
The internal syntax of numeral 1000 in Pattern 2 would have the following form:

(35)  
\[ \text{\#} \quad \#_{\text{PL}} \quad \sqrt{\text{P}} \quad \sqrt{1000} \quad \ldots \]

The use of quantifiers in the extended projection allows for the creation of complex numerals (see e.g. Ionin and Matushanksy 2006). With numeral 1000, the syntax is fully transparent, meaning that the case and number inflection of 1000 is exactly what a masculine inanimate noun would show in the given context.3

(38) Adjectives & quantifiers in the extended projection:
   a. Tych / kilka / kilkaś niedznych tysiąc dolców
       DEM.PL GEN tens(11-19) miserável.PL GEN thousands GEN dollars GEN
       ustawił/ó ich finansowo …
   b. Kilka / kolejnych /tyścicy /ptaków
       Several //many/how,many thousands GEN birds GEN
       ‘Several thousand birds’ / (how) many thousands of birds’
   c. Kilka / wiele / ile /tyścicy / ptaków
       Several / many / /how, many thousands GEN birds GEN
       ‘Several thousand birds’ / (how) many thousands of birds’

The presence of an extended projection which hosts such material is further evidence of the nominal status of the numeral (with and without gender).

We expect (approximately) the following syntax for the numeral (assuming for now, adjectives as in (38) appear lower than a QP, e.g. in Spec,#P):

3 Though pre-modifiers are capable of targeting the noun, agreeing numerals like 1 or 2 never do, e.g.: (i) Jednych / tysiace listów / ‘one thousand letters’ / two.PL GEN / three.PL GEN / four.PL GEN 1000s letters / DePL GEN 1000 letters

   ‘One thousand letters, two/three/four thousand letters, those thousand letters’
3.2.2 External syntax of Polish 1000

The basic word order of a Polish nominal expression involves: demonstrative, quantifier, adjective, noun (the base word order of DPs according to Cinque 2005, Abels and Neeleman 2006).

This suggests that Polish numerals have a dedicated region in the nominal extended projection, which I will call QP (following e.g. Löbel 1989).

Whether numerals belong in the specifier or head of QP is a question with a complicated (and at times theory-dependent) history. Willim (2015) cites left branch extraction as motivation for a specifier analysis (assuming only phrases extract):

Regardless of the precise analysis, it’s clear that the numeral is located in the quantificational region of the extended projection of the quantified noun. This can be added to the lexical specification of the numeral, by assigning it a feature “Q”:

Case assignment and agreement (formally):

I’ve adopted a feature sharing (Frampton and Gutmann 2000; Pesetsky and Torrego 2007, a.o.) and dependent case approach (Marantz 1991; Baker 2015; Levin 2015, a.o.) (see Klockmann 2017 for discussion).

Pattern 1, full agreement:

• Numerical and noun both associated with phi-complete phi-bundles (φ₁ and φ₂).
• φ₁ (numeral) functions as an intervener, blocking agreement with φ₂ (noun).
• Agreement by all probes is with φ₁ (numeral).

Evidence from agreement in a complex numeral is suggestive of a specifier analysis. Taking a lower numeral (2–4), which usually agrees in gender with what it quantifies, and crucially here, does not trigger case, an adjective external to the complex numeral appears to surface in either the default form or in agreement with the quantified noun (i). An adjective internal to the complex numeral, however, seems unable to agree with the noun (ii). This evidence is preliminary.

(i) Kolejne / *kolejnych trzy tysiące złotych

(ii) Trzy kolejne / *kolejnych tysiące złotych

‘the next three thousand’
3.3 Structural vs. oblique case contexts

Recall the case properties of Polish numeral 1000 – in structural and oblique case contexts, the quantified noun appears in the genitive case.

(47) a. Tysiąd ptaków  
    Thousand.NOM/ACC  birds.Gen  
    ‘Thousand birds’

b. Z tysiędzicem ptaków  
    With[ILLUS] thousand.INST  birds.Gen  
    ‘With a thousand birds’

It can be shown that only the fully lexical version of numeral 1000 is permitted in an oblique case context. Consider the form of the pre-modifier:

(48) a. Z tym/*tym/*te tysiącem ludzi  
    ‘With those thousand people’

Reasoning:
- The default phi-feature values of Polish pre-modifiers is non-virile plural (22).
- The DP the numeral is contained in will be constructed before it is merged with the instrumental-requiring preposition z.
- If agreement is with the numeral, which lacks γ, then agreement should default.
- The addition of instrumental case is not predicted to interact with prior # and γ agreement.
- Therefore, we predict a default feature specification on the pre-modifier, e.g. NV.PL. This is not possible! Neither as INST or as NOM.
- Only masculine singular features are permitted on the pre-modifier.
- This suggests the numeral carries γ in an oblique environment.

Finding: The semi-lexicality of the numeral plays a role in what contexts it can appear in. Semi-lexical 1000 is not permitted in an oblique environment, and instead, lexical 1000 must occur.

Comparison to 5+ numerals:

The genitive of 5+ numerals (5-10, 100) is lost in an oblique environment. 5+ numerals have no lexical equivalent, unlike numeral 1000.
4. English Q-nouns

English quantificational nouns (Q-nouns) share the property that they appear to have the morphosyntax of a noun, despite functioning as a quantifier:

(51) a. A lot of books, lots of books
    b. A ton of books, tons of books
    c. A bunch of books

For *ton* and *bunch*, the quantificational interpretation competes with a lexical interpretation in the morphosyntax of (51).

(52) a. A ton of bricks (= one ton [2000 pounds] of bricks, OR a lot of bricks)
    b. A bunch of herbs (= herbs picked and bundled as a bunch “literal *bunch*”, OR a lot of herbs)

*a* There were no genuine examples involving a pre-modifier, and while a native speaker was found who accepted (50), the addition of a pre-modifier necessitated the pattern of (47), i.e. genitive on the quantified noun. This makes it difficult to use agreement to ensure this is semi-lexical 1000, though given the analysis so far, it is predicted to be.

Lot also has a competing collection interpretation, though it seems to be absent in the morphosyntax of (51):

(53) a. Old maids, the lot of you (COCA: Bk:IrishBlood 2015)
    b. I would prefer to damn this lot of un忍受able bores to the netherworld (COCA: Bk:MyLordVampire 2012)
    c. A lot of un忍受able bores (= many un忍受able bores?)

While I only report on *lot*, *ton*, and *bunch* here, the morphosyntax of (51) appears to represent a productive way of creating new quantificational expressions. The examples below, extracted from the COCA all seem to mean something like “many”, with imagery form the lexical semantics of the Q-noun coloring the interpretation:

(54) A flood of memories, a parade of witnesses, a torrent of words, a sea of faces, a flurry of lawsuits, a mass of contradictions, a chorus of boos, a mob of reporters, a multiplicity of voices, a barrage of insults, a host of possibilities, a cascade of problems

The findings reported below are based on native speaker judgments, and the Corpus of Contemporary American English (Davies 2008), marked as (COCA: Source, Year).

4.1 Idiosyncrasies of Q-nouns *lot*, *ton*, *bunch*

The quantificational nouns (Q-nouns), *lot*, *ton*, and *bunch*, cannot function as agreement targets, despite their apparent singular or plural morphosyntax:

(55) a. A lot of people were / *were missing.
    b. Lots of sugar was / *were spilled.
    c. A ton of people were / *were missing.
    d. Tons of sugar was / *were spilled.
    e. A bunch of people were / *were missing.

However, they show minor idiosyncratic differences, which suggest there is no one fits-all analysis possible. Instead, I propose, they have a similar syntax, but it must adhere to the lexical requirements of individual Q-nouns.

4.1.1 Plurality

*Lot* and *ton* can be morphologically plural, while *bunch* cannot.
The inability to surface in a plural form is a property of Q-noun bunch, as lexical bunch does permit plural marking:

(57) Literal bunch:
   a. She hung more crystals and bunches of herbs throughout the room. (COCA: LiteraryRev 1990)
   b. Angelo purchased two bunches of cut flowers. (COCA: Bk:Blindsight 1992)

Though the missing plurality is subject to variation, as Q-noun examples surface:

(58) a. I don’t have bunches of time left. (COCA: CBS_Morning 1992)
   b. Well, I’ve had bunches of different kinds of memory loss. (COCA: NPR 2013)

The plurality on lot and ton does not seem to indicate semantic plurality. Numerals and quantifiers cannot quantify these plural forms:

(59) a. *One lot of mistakes
   b. *Two / many / a few lots of mistakes
   c. *One ton of mistakes
   d. *Two / many / a few tons of mistakes

Together, this shows that Q-nouns differ in their ability to inflect as plural. We can model this in the lexical entries as follows:

Lexical entries of lot, ton, bunch (to be revised):

- Lot: [√lot, Q, INDEF] Root specified for indefinite contexts
- Ton: [√ton, Q]
- Bunch: [√bunch, Q, ¬#pl] (root specified for absence of plural morphology)

4.1.2 Definiteness

Ton can appear with definite DP material, while lot and bunch cannot.

(60) Ton, tons:
   a. I quite frankly was very surprised that despite the ton of hype and all the
   excitement over Howard’s debut, the ratings weren’t much bigger than they were. (COCA: CNN_Showbiz 2012)
   b. Had I looked into a crystal ball and seen the tons of extra stress this would cause, well, frustration is getting the upper hand now. (COCA: AssocPress 2007)
   c. I don’t know who helped this fellow take out this ton of junk food. (COCA: SatEvenPost 2008)
   d. the “hundreds of thousands of people” who wrote to CBS and all those tons of nuts… (COCA: WashPost 2007)

(61) Lot, lots (under the Q-noun reading):
   a. *The lot of people
   b. *The lots of butter
   c. *This / that lot of people
   d. *These / those lots of butter

(62) Bunch (under Q-noun reading)
   a. *The bunch of people
   b. *This/those bunch of people

We can embellish the lexical entries of the Q-nouns to indicate their ability to occur in a definite context. Two options are possible here: INDEF or ¬DEF.

Revised lexical entries:

- Lot: [√lot, Q, INDEF] Root specified for indefinite contexts
- Ton: [√ton, Q]
- Bunch: [√bunch, Q, INDEF, ¬#pl] Root specified for indefinite contexts

4.2 The morphosyntax of Q-nouns

Observation: With productive cases of Q-noun formation, an indefinite article and of necessarily appears:

(63) a. *A wealth
   b. *Wealth of money
   c. A wealth of examples (= ‘many examples which are rich in nature’)

→ The indefinite and particle of are properties of the construction.
4.2.1 The article

**Proposal:** The article is a lexicalization of Q, which surfaces because the Q-noun cannot move to Q.

\[(64) \quad \text{QP} \]

\[Q \quad \sqrt{P} \quad \sqrt{\text{Q-noun}} \ldots \]

**Motivation:** The article surfaces in a number of quantificational expressions, where it does not control agreement (65). Furthermore, its presence can be triggered by the insertion of an adjective on a quantificational expression (66).

\[(65) \]

a. A hundred birds were found nesting.
b. A couple birds were found nesting.
c. A dozen birds were found nesting.
d. A lot / ton / bunch of birds were found nesting.

\[(66) \]

a. A great many birds were found nesting.
b. A mere eight companies own nearly all of mainstream media journalism (COCA: USA Today 2007)
c. A suffocating 92,000 fans were in the seats for each game at the L.A. Memorial Coliseum. (COCA: Chicago 2005)
d. Bypassing or tampering with power supplier meters is a growing problem, costing an estimated hundreds of billions of dollars worldwide and… (COCA: Futurist 2013)

Supposing the presence of the adjective blocks lexicalization of Q with numerals and many, the article is triggered:

\[(67) \]

**Lexicalization of Q with and without modifier**

a. QP

\[Q \quad \sqrt{P} \quad \sqrt{\text{three}} \ldots \]

b. QP

\[Q \quad \sqrt{AP} \quad \sqrt{P} \quad \sqrt{\text{three}} \ldots \]

**Claim:** The article is a “dummy” article used as a last-resort to fill an empty Q (see Lyons 1999 for a similar analysis of the article a(n) in English).6

4.2.2 Plurality and definiteness with Q-nouns

The lexical specifications of Q-nouns interact with this basic syntactic configuration, generating idiosyncrasies. I repeat the lexical entries here:

- Lot: \([\sqrt{lot, Q, \text{INDEF}}]\)
- Ton: \([\sqrt{ton, Q}]\)
- Bunch: \([\sqrt{bunch, Q, \text{INDEF}, \neg\#_{pl}}]\)

**Plurality:** Both lot and ton allow plurality. But, the plural marker does not trigger the default article:

\[(68) \]

a. *A lots of books
b. *A tons of books

This suggests Q is lexicalized, likely by the apparent plurality marker, which spells-out on the Q-noun:

\[(69) \]

**Implication:** The \([\neg\#_{pl}]\) specification on bunch is a condition on which Q’s it can combine with.

**Definiteness:** Neither lot nor bunch allow definiteness, a condition on the realization of DP. There are two plausible analyses: (a) if indefiniteness is the absence of definiteness (Lyons 1999), these Q-nouns are incompatible with the DP layer, or (b) the Q-nouns are incompatible with a definite specification on D.

---

6 This immediately predicts the sequence "the a" if the QP is dominated by a DP. Lyons (1999) suggests this is blocked by a phonological constraint; Boeber (2005: 160-169) faces the same issue, and proposes to analyze the as originating lower in the DP and moving up to D (this blocking the presence of a).
5. Dutch generic soort

5.1 Soort according to The syntax of Dutch

(71) Two types of 'soort' (excluding an approximator reading)

a. Deze / die soort aap/aapen Soort 1
   This.C / that.C kind monkey / monkey.PL

b. Dit / dat soort auto/auto's Soort 2
   This.N / that.N kind car / car.PL

(Broekhuis & den Dikken 2015: 631, ex. 158a, b)

71a: “...the noun soort is clearly used as a referential expression and the binominal construction refers to a contextually determined species of monkey.” (Broekhuis & den Dikken 2015: 631)

Properties of soort 1 (according to Broekhuis & den Dikken 2015):
- Has common gender, cannot co-occur with an N2 which is neuter, pluralizes
- Verbal agreement targets soort 1

71b: “...has a type reading in the sense that it refers to a set of cars that resemble a certain car/certain cars that is/are under discussion” (Broekhuis & den Dikken 2015: 631)

Properties of soort 2 (according to Broekhuis & den Dikken 2015):
- Has neuter gender, N2 not restricted in gender, no plural?
- Verbal agreement targets soort 2 or N2.

5.2 Findings, via corpus & native speaker discussion

Corpus: Corpus of Contemporary Dutch (Corpus Hedendaags Nederlands), (CoCD), using the subcorpus of “Dutch Dutch” and a subset of 1000 examples

Two interpretations of soort, roughly:
(a) Established kind, e.g. via world knowledge (e.g. animal species) or context
(b) Out-of-the-blue kind, determined on the spot via an exemplar

Working hypothesis: These show different agreement properties:
(a) Established kind: Demonstrative/adjective agreement with noun
Verbal agreement with soort or noun
(b) Out-of-the-blue kind: Demonstrative/adjective agreement with soort
Verbal agreement with noun or soort (formal)

Let’s call them soort 1 (established kind?) and soort 2 (out-of-the-blue kind?).
5.2.2 DP-internal agreement: Determiners, demonstratives & adjectives

**Generalization:** DP-internal modifiers (determiners, demonstratives, and adjectives) agree with the noun for soort1 and with soort (as neuter) for soort2.

**Manipulation:** Mismatch in gender (neuter (N) vs. common (C), the (a) examples) or mismatch in number (neuter singular (N) vs. plural (PL), the (b) examples)

**Determiners**

(75) **Agreement with noun** (soort1: established kind?) [CoCD]

a. De soort spanning die voetballers en supporters…
   The C kind excitement that football players and supporters
   ‘The kind of excitement that football players and supporters…’

b. Te beginnen met de officiële promotiecampagne, zoals …
   To begin with the official promotion campaign such as
   ‘To begin with the official promotional campaign, such as… giving
   information about the kinds of houses that…’

(76) **Agreement with ‘soort’** (soort2: out-of-the-blue kind?) [CoCD]

a. Het soort regen dat uit de zee wordt aangedragen
   The N kind rain that from the sea is brought
   ‘The kind of rain that is brought out of the sea by low dark clouds that…’

b. Hij hoorde niet tot het soort componisten dat zegt: doe maar
   He belonged not to the N kind composers that says: do just
   ‘He did not belong to the kind of composers that say: do what you want
   so long as it sounds convincing.’

**Demonstratives**

(77) **Agreement with noun** (soort1: established kind?) [CoCD]

a. Dit en vorig jaar werd deze soort belasting met een stuiver
   This and last year was this C kind tax with a five-cent
   raised
   ‘This year and last year, this kind of tax was increased by five cents’

b. Deze soort tegels worden op vele plaatsen toegepast en vormen
   These kind tiles are on many places fit/used and form
   geen bellemmeringen
   no obstacles.
   ‘These kind of tiles are used in many places and do not form an obstacle.’

(78) **Agreement with ‘soort’** (soort2: out-of-the-blue kind?) [CoCD]

   For this N kind employment, C sits Drenthe to wait
   ‘Drenthe is waiting on this kind of employment.’

b. Dat soort bedrijven waren er nodig om de
   That N kind companies were there needed in order to
   Europese landbouwpolitiek uit te voeren
   European agricultural policy to run
   ‘Those kind of companies were needed in order to implement
   the European agricultural policy.’

**Attributive adjectives**

(79) **Agreement with noun** (soort1: established kind?) [CoCD]

a. Niet alleen de fietser op trektocht of een andere soort tourist
   Not only the biker on hiking tour or another C kind tourist, C
   komt voor huur in aanmerking
   ‘Not only the biker on tour or another kind of tourist is eligible for rent.’

b. Op latere leeftijd gaan ze stelen om lijm te kunnen kopen,
   At later ages go they steal to glue to can buy
   de goedkoopste soort drugs
   the PL cheapest PL kind drugs PL
   ‘At a later age, they steal to be able to buy glue, the cheapest kind of
   drugs’

(80) **Agreement with ‘soort’** (soort2: out-of-the-blue kind?) [CoCD]

a. Dat is een heel ander soort angst
   That is a whole other N kind fear C
   ‘That is a whole other kind of fear.’

b. …illegal groeperingen, …, de opdracht kregen uit te kijken
   illegal groups … the assignment received to look
   naar een nieuw soort jachtvliegtuigen dat op deze basis
   for a new kind jet fighter that on this base
   gestationeerd zou worden.
   stationed shall be
   ‘Illegal groups received the assignment to look out for a new kind of
   fighter jet what would be stationed on this base.’
Two agreement patterns: (a) agreement with the noun (soort1), or (b) agreement with soort as neuter (soort2).

5.2.3 DP-external agreement: Verbs

Generalization: Verbal agreement targets noun, and sometimes soort2?
Manipulation: Mismatch in number (singular soort vs. plural noun)

5.2.4 Plural soorten ‘kinds’

Plural soorten:
- The out-of-the-blue kind interpretation seems to be missing with soorten
- Pre-modifiers always agree with soorten
- Verbs always agree with soorten

5.2.5 Summary & preliminary conclusions

- Dutch appears to have grammaticalized the “established-ness” of the subkind, through varying the agreement properties? → suggests there is a grey area in between where both types of agreement patterns may occur.
- Soort1: “Established kind” – DP-internal agreement with noun, verbal agreement with noun or soort
• Soort2: “Out-of-the-blue kind” – DP-internal agreement with soort (neuter), verbal agreement with noun or soort (formal)

• Soorten: Always controls agreement

5.3 Preliminary analysis

I set verbal agreement aside for the moment, pending a clearer picture on what conditions the verbal agreement patterns.

Instead, let’s explore the following paradigm:

• Soort1: Agreement targets the noun, surfacing in the features of the noun

• Soort2: Agreement targets soort, surfacing as neuter

• Soorten = pluralization of soort1?

This resembles the pattern of agreement found with Polish 1000:

(86) Tych /te tysiąc ptaków
DEM.GEN /NV.PL.NOM/ACC thousand birds.GEN
‘Those thousand birds’

A similar analysis is applicable, assuming:

• Soort1 carries number but not gender (and hence can pluralize)

• Soort2 carries gender but not number (and hence cannot pluralize)

• DP internal agreement “prefers” to agree with something specified for γ (unless it finds something plural).7

• (Potentially: DP external agreement “prefers” to agree with something specified for #, or to adopt a prescriptive norm of agreeing with soort)

(87) a. [DP D [SoortP soort1 [NP N [γ, #] [γ, #]]]

b. [DP D [SoortP soort2 [NP N [γ, #] [γ, #]]]

7 This type of phenomenon sounds like context sensitive agreement, where agreement probes are sensitive to the feature values of a goal (see Bejar 2001). Potentially, under an approach like Hachem (2015), Dutch has gender and plural number, but nothing called “singular” per se.

6. Conclusions

• Not everything fits neatly into the lexical, functional dichotomy, hence the recurrence of the descriptive term “semi-lexical”.

• I proposed that semi-lexicality is what arises when a root is specified for functional features, those features placing requirements on the syntax.

• This was illustrated with three case studies: Polish 1000, English Q-nouns, and Dutch generic soort

• Polish 1000 was proposed to have the syntax of a noun, though with the absence of gender, this having effects on agreement. This is modeled in the lexical entry.

• English Q-nouns were argued to have a similar morphosyntax, but differ in their distributional requirements, some allowing apparent plurality and others definiteness. This is modeled in the lexical entry.

• Dutch soort was shown to have two agreement patterns, which could be accommodated by making assumptions on the features carried by soort.

7. References


