The Danish Morphosyntactically Tagged PAROLE Corpus

1. Introduction

As shown in the table below, the Danish PAROLE tagset does not include determiners, articles and numerals. Instead these words are covered by other parts-of-speech. The language-specific (i.e. Danish) features in positions 8-11 in the tagset table below include: (i) **definiteness** (definite/indefinite) for nouns and adjectives, (ii) **voice** (active/passive) for some verb forms, (iii) **reflexiveness** (yes/no) for some personal and possessive pronouns, (iv) **register** (formal/obsolete/polite/unmarked for register) for most pronouns, as well as (v) **transcategorization** (adjectival/adverbial/unmarked for usage) for adjectives and participles. The latter necessitates the presence of other adjectival features in the description of participles that are used adjectivally (such as number, gender, **definiteness**, and **case**).

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Noun		Gender	Number	Case			Definiteness			
Verb		Mood	Tense	Person	Number	Gender	Definiteness	TrCat	Voice	Case
Adj		Degree	Gender	Number	Case		Definiteness	TrCat		
Pron		Person	Gender	Number	Case	Possessor	Reflexive	Register		
Det		Person	Gender	Number	Case	Possessor				
Art		Gender	Number	Case						
Adv		Degree	Function	Wh-ness						
Adpos		Formation	Gender	Number						
Conj		Ctype	Coord-posit							
Num		Gender	Number	Case						
Interj										
Residual										
Unique										

2. Parts-of-Speech

This chapter briefly reviews the different parts-of-speech that appear in the Danish PAROLE tagset above. A more thorough presentation of the decisions made for each of the different parts-of-speech is found in the full Danish documentation for the morphosyntactically tagged subcorpus, 'Vejledning til det danske morfosyntaktisk taggede PAROLE-korpus'.

2.1 Nouns

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Noun N	Common C	Gender	Number	Case	=	=	Definiteness			
Noun N	Proper P	-	-	Case	=	=	-			

As shown in the table above, Danish common nouns ('NC') are marked for inherent gender (common/neuter), as well as inflection in number

(singular/plural), case (genitive case/unmarked for case, i.e. non- genitive), and definiteness (definite/indefinite). Danish proper nouns ('NP') are only marked for case (genitive/unmarked for case). In general all nouns that overlap with regular common nouns are tagged as common nouns, regardless of whether they are used (and hence capitalised) as proper nouns, either alone or as part of multiword naming expressions. Hence, *Statsministeren* ('The Prime Minister') and *Teater* in *Det kgl. Teater* ('The Royal Theatre') are both tagged as common nouns, and not as proper nouns. There are, however, a few exceptions to this general rule, as will be briefly outlined below.

..MAN må kritisere det kuldslåede <u>fornuftsægteskab</u> med medierne og branchen..
    <W lemma="fornuftsægteskab" msd="NCNSU==I">fornuftsægteskab</W>
..to fluers indædte kamp om <u>majsgrødresterne</u> på en uslikket ske og fars hivende åndedræt larmede..
    <W lemma="majsgrødrest" msd="NCCPU==D">majsgrødresterne</W>
..i debatten tordnes der løs mod Det kgl. <u>Teaters</u> repertoire..
    <W lemma="teater" msd="NCNSG==I">Teaters</W>
..samt tre af de nye Standardflex 300 glasfiber- patruljefartøjer '<u>Havkatten</u> ", "<u>Hajen</u> " og "<u>Makrellen</u> "..
    <W lemma="havkat" msd="NCCSU==D">Havkatten</W>
    <W lemma="haj" msd="NCCSU==D">Hajen</W>
    <W lemma="makrel" msd="NCCSU==D">Makrellen</W>
..resterne af Jyske <u>Division</u> skal indgå i NATOs <u>Hovedforsvarstyrke</u>, hedder det i rapporten..
    <W lemma="division" msd="NCCSU==I">Division</W>
    <W lemma="hovedforsvarsstyrke" msd="NCCSU==I">Hovedforsvarsstyrke</W>

When the name of a person or a geographical location overlaps with a (theoretically possible) common noun, the word has still been tagged as a proper noun.

..det kan kun give en masse rutine," siger amatørspilleren Thomas <u>Bjørn</u>..
    <W lemma="Bjørn" msd="NP- - U==- ">Bjørn</W>
..sammen med <u>Blåbjerg</u> Kommunes borgmester .. tager Arne Toft .. til Christiansborg..
    <W lemma="Blåbjerg" msd="NP- - U==- ">Blåbjerg</W>

A few Danish names of geographical locations are inflected in number and/or definiteness, as for example *Færøerne* ('The Faroe Islands') or *Elfenbenskysten* ('The Ivory Coast'). However, they are still tagged as proper nouns despite their common noun inflections. A few proper nouns, as for example the names of products, occasionally appear in inflected forms in the corpus. In these rare cases they have been tagged as common nouns on the basis of their nominal inflections.

.."Men <u>Færøerne</u> tilhører dog det danske rigsfælleskab," siger Erik Hjorth Nielsen..
    <W lemma="Færøerne" msd="NP- - U==- ">Færøerne</W>
..på vej mod Bregenz glæder vi os endnu til turen til <u>Alperne</u>..
    <W lemma="Alperne" msd="NP- - U==- ">Alperne</W>
..han havde kurs direkte mod <u>Fiat'en</u>, og jeg så føreren af Fiat'en vride en gang i rattet..
    <W lemma="Fiat" msd="NCCSU==D">Fiat'en</W>
..de sidste par nætter har nazisterne brugt parkerede <u>Trabanter</u> til at barrikadere gaderne..
    <W lemma="Trabant" msd="NCCPU==I">Trabanter</W>

For compound nouns, the exact distinction between nouns denoting a geographical location and common nouns is more difficult to define, but compound nouns that denote names of geographical locations have always been tagged as proper nouns, regardless of possible common noun inflections. Some examples of this are: *Fælledparken, Halmtorvet, Hareskoven, Holbækmotorvejen, Kvægtorvet, Lillebæltsbroen, Limfjordtunnelen, Lyngbyvejen, Rådhuspladsen,* and *Sjællandsbroen .* When appearing in multiword naming expressions denoting geographical locations, a very limited number of common nouns (and occasionally adjectives) has also been tagged as proper nouns. Capitalised acronyms are tagged as proper nouns unless they are listed in the official Danish Orthographical Dictionary (Dansk Sprognævn, 1996) as common nouns (in which case they need not be capitalised).

> **..satspuljen stammer fra et tidligere forlig om overførselsindkomster, hvor også <u>SF</u> deltog..**
>   <W lemma="SF" msd="NP- - U==- ">SF</W>
> **..resterne af Jyske Division skal indgå i <u>NATOs</u> Hovedforsvarstyrke, hedder det i rapporten..**
>   <W lemma="NATO" msd="NP- - G==- ">NATOs</W>
> **..mottoet er: "Ka' du læse billeder, ka' du også lære <u>EDB</u>."..**
>   <W lemma="edb" msd="NCCSU==I">EDB</W>
> **..konkurrencen kommer af, at flere vil udvikle ny teknologi, informationssystemer, <u>edb</u>,..**
>   <W lemma="edb" msd="NCCSU==I">edb</W>

In Danish, it is not uncommon for the head noun not to be expressed in a noun phrase, especially when referring to people, e.g. *de gamle* ('the old/the elderly [people]'), or *en arbejdsløs* ('an unemployed [person]'). This is sometimes referred to as the 'nominal usage' of the last adjective/participle/numeral in the noun phrase (cf. Allan *et al,* 1995). However, the special status of these adjectives, etc. has not been made explicit in the PAROLE corpus, and they are tagged in the same way as all other adjectives, etc.

> **..det var indtil denne uges meninghedsrådsvalg. Nu står det syv- seks. Og det er altså i de <u>helliges</u> disfavør..**
>   <W lemma="hellig" msd="ANP[CN]PG=[DI]- ">helliges</W>
> **..den dag, Sansolo talte som den <u>tolvte</u> i ørkenen, slap vandet op. Kamelerne gav mindre og mindre mælk..**
>   <W lemma="tolvte" msd="AO- - - U=- - ">tolvte</W>
> **..en journalist, der bevægede sig igennem ruinerne i Erzincan efter <u>overlevende</u> ..**
>   <W lemma="overleve" msd="VAPR=[SP][CN][DI]A- U">overlevende</W>
> **..“Jeg har hele tiden kuldegysninger ned ad ryggen. Jeg frygter, der er mange <u>dræbte</u> ," sagde en talsmand..**
>   <W lemma="dræbe" msd="VAPA=P[CN][DI]A- U">dræbte</W>

Foreign loanwords in Danish which are listed in dictionaries such as Dansk Sprognævn, 1996 (e.g. *at booke, at sample, cool, macho, en cheddar, en D- mark, en fight,* etc.) are naturally tagged on par with other Danish words. However, words of a foreign origin belonging to the following three categories were not listed in Danish dictionaries and hence needed special attention in the PAROLE corpus: (i) foreign words in foreign quotations (*Havde Goethe ikke sagt: Dort wo du nicht bist, ist das Glück.*), (ii) foreign words that are the names of Danish or foreign institutions, firms or products (e.g. *Accumulator Invest, Coca- Cola,*

Danisco, Financial Times, Holiday Inn, Kronfågel, Macintosh, Standardflex, Toyota, WordPerfect), names of music groups/bands (e.g. *Status Quo, Rolling Stones, Nick Cave and the Bad Seeds*) or and (iii) foreign words in foreign book/record/song/film titles (e.g. *The Natural History of Alcoholism*, *Star Wars, Monsieur Hire*). After an examination of the morphological inflections and syntactic functions of these words in the PAROLE corpus, it was decided to tag these foreign words as proper nouns ('NP') if they are capitalised in the corpus, and as 'foreign words' ('XF') if they are not.

..en af de begejstrede kommentarer kommer fra <u>Kvällspostens</u> sportskommentator..
    <W lemma="Kvällsposten" msd="NP- - G==- ">Kvällspostens</W>
..Kræftens Bekæmpelse synes at opfatte bidragyderne .. som en <u>Coca- Cola</u> frabrikant betragter forbrugerne..
    <W lemma="Coca- Cola" msd="NP- - U==- ">Coca- Cola</W>
..en salat tilberedt af fiskerogn, <u>tacik</u> (yoghurt med agurk, gulerod og hvidløg), <u>dolma</u> (marinerede vinblade)..
    <W lemma="tacik" msd="XF">tacik</W>
    <W lemma="dolma" msd="XF">dolma</W>
..tilbage til den lille hyggelige østriske alpeby med <u>gemütlichkeit</u>, store værelser og ditto senge..
    <W lemma="gemütligchkeit" msd="XF">gemütligchkeit</W>

2.2 Verbs

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Verb V	**Main** A	**Mood**	**Tense**	**=**	**Number**	**Gender**	**Definiteness**	**TrCat**	**Voice**	**Case**
Verb V	**Medial** E	**Mood**	**Tense**	**=**	**Number**	**Gender**	**Definiteness**	**TrCat**	**Voice**	**Case**

The Danish verbs are divided into 'main' ('VA') and 'medial' ('VE') verbs, where the latter category encompasses deponent and reciprocal verbs, which are passive in form, yet active in meaning (cf. Allan *et al,* 1995). Moreover, verbs are classified according to mood as indicative, infinitive, imperative, gerund, and participle verbs. Modal verbs and auxiliary verbs are not tagged explicitly in the Danish PAROLE corpus. As can be seen from the table above, Danish verbs are marked for inflection in tense, number, gender, definiteness, transcategorization, voice, and case. However, only present and past participles are marked for the adjectival inflections (number, gender, definiteness, transcategorization and case). Separable complex verbs, reflexive verbs, and other multiword verbal expressions are not explicitly linked in the Danish PAROLE corpus, and instead each word is tagged independently.

..Anders Bircow, 40, <u>slapper</u> i denne tid <u>af</u> efter en succésfuld sæson..
    <W lemma="slappe" msd="VADR=- - - - A- ">slapper</W>
    <W lemma="af" msd="RGU">af</W>
..til sidst <u>stod</u> hun dog <u>op</u> og fik så øje på et uhyggeligt orangegult skær..
    <W lemma="stå" msd="VADA=- - - - A- ">stod</W>
    <W lemma="op" msd="RGU">op</W>
..de lo alle sammen. Så <u>skyndte</u> de <u>sig</u> at samle tallerkenerne sammen, så de kunne komme af sted..
    <W lemma="skynde" msd="VADA=- - - - A- ">skyndte</W>
    <W lemma="sig" msd="PP3[CN][SP]U- YU">sig</W>

..i forkontoret er vi en ryger og en ikke- ryger. Vi må jo se at <u>finde</u> <u>ud</u> <u>af</u> det..
    <W lemma="finde" msd="VAF- =- - - - A- ">finde</W>
    <W lemma="ud" msd="RGU">ud</W>
    <W lemma="af" msd="SP">af</W>

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Verb V	Main A	Indicative D	Present R	=	-	-	-	-	Active A	-
Verb V	Main A	Indicative D	Present R	=	-	-	-	-	Passive P	-
Verb V	Main A	Indicative D	Past A	=	-	-	-	-	Active A	-
Verb V	Main A	Indicative D	Past A	=	-	-	-	-	Passive P	-
Verb V	Medial E	Indicative D	Present R	=	-	-	-	-	Active A	-
Verb V	Medial E	Indicative D	Past A	=	-	-	-	-	Active A	-
Verb V	Main A	Imperative M	-	=	-	-	-	-	-	-
Verb V	Main A	Infinitive F	-	=	-	-	-	-	Active A	-
Verb V	Main A	Infinitive F	-	=	-	-	-	-	Passive P	-
Verb V	Medial E	Infinitive F	-	=	-	-	-	-	Active A	-

As shown in the table above, the indicative verb forms ('VAD' or 'VED') are marked for inflections in tense and voice, while the imperative verb form ('VAM') is unmarked, and the infinitive verb forms ('VAF' or 'VEF') are only marked for inflection in voice.

..måske fordi lakkjolen var så stram, at den knirkede, når hun <u>gestikulerede</u> for meget..
    <W lemma="gestikulere" msd="VADA=- - - - A- ">gestikulerede</W>
..den blå røg fra knallerterne <u>spredtes</u> , og drønet fra bollespisernes motorer fortabte sig i det fjerne..
    <W lemma="sprede" msd="VADA=- - - - P- ">spredtes</W>
..<u>spring</u> rundt og <u>dans</u> til radioen, <u>køb</u> en børnepose hos slikmutter, og <u>æd</u> den selv..
    <W lemma="springe" msd="VAM- =- - - - - ">Spring</W> ... <W lemma="danse" msd="VAM- =- - - - - ">dans</W>
    <W lemma="købe" msd="VAM- =- - - - - ">køb</W> ... <W lemma="æde" msd="VAM- =- - - - - ">æd</W>
..men det vigtigste er ikke at <u>lade</u> sig smitte af depressions- bacillen!..
    <W lemma="lade" msd="VAF- =- - - - A- ">lade</W>

The expression 'gerund' ('VAG') is used to cover verbal nouns of the type shown below:

..kunst og kunstnerisk <u>skaben</u> handler om modsætninger og deres frugtbare interaktion..
    <W lemma="skabe" msd="VAG- =SCI- - U">skaben</W>
..ingen andre kan formentlig få Bob Dylan til at sidde og iagttage sig med drenget <u>undren</u> ..
    <W lemma="undre" msd="VAG- =SCI- - U">undren</W>

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Verb V	Main A	Participle P	Present R	=	Number	Gender	Definiteness	TrCat	-	Case
Verb V	Main A	Participle P	Past A	=	Number	Gender	Definiteness	TrCat	-	Case

CatGram m Verb V	SsCatGra m Medial E	3 Participle P	4 Past A	5 =	6 Number	7 Gende r	8 Definitene ss	9 TrCat	10 -	11 Case

As shown in the table above, present participles ('VAPR') and past participles ('VAPA') are marked for tense (present/past), number (singular/plural), gender (common/neuter), definiteness (definite/indefinite), transcategorization (adjectival/adverbial/unmarked for usage) and case (genitive/unmarked for case). The inflections for number, gender, definiteness, transcategorization, and case are only relevant for participles that function adjectivally. In Danish, present and past participles can be used either verbally and adjectivally (and hence occasionally also adverbially). However, in most cases <u>no</u> distinction has been made in the tagging of participles used verbally or adjectivally. Hence participles are mostly underspecified for transcategorization in the Danish PAROLE corpus (with 'ARU' signifying either adjectival, adverbial, or unmarked - i.e. verbal - usage). However, there are three cases in which participles are <u>not</u> underspecified for usage in this way: (i) when they appear attributively in a noun phrase, (ii) when they appear (anywhere) in an inflected form that does not overlap with the verbal form, and (iii) when they appear attributively in an adjectival or adverbial phrase.

..sociallovgivningen, der omfattede Forsorgsloven - <u>svarende</u> til den nuværende Bistandslov..
        `<W lemma="svare" msd="VAPR=[SP][CN][DI][ARU]-U">svarende</W>`
..jeg er <u>kommet</u> til at hade samfundet og øvrigheden for den råddenskab, jeg har været udsat for..
        `<W lemma="komme" msd="VAPA=S[CN]I[ARU]-U">kommet</W>`
 ..det var ikke et sted, man lod sig se, hvis man ville være velset i byens <u>herskende</u> indremissionske kredse..
        `<W lemma="herske" msd="VAPR=[SP][CN][DI]A-U">herskende</W>`
..alkoholpromiller mellem 2,00 og 2,50 er takseret til den <u>nævnte</u> straf..
        `<W lemma="nævne" msd="VAPA=S[CN]DA-U">nævnte</W>`
..injektioner med genbrugte engangskanyler eller kanyler, der ikke er <u>rensede</u> tilstrækkeligt..
        `<W lemma="rense" msd="VAPA=P[CN][DI]A-U">rensede</W>`
..hun har samlet en stor flok unge, danseglade og <u>overraskende</u> dygtige balletfolk..
        `<W lemma="overraske" msd="VAPR=- - -R-U">overraskende</W>`
..en undersøgelse af ernærings- og sundhedstilstanden blandt en .. gruppe, <u>formodet</u> raske Østeuropæiske børn..
        `<W lemma="formode" msd="VAPA=- - -R-U">formodet</W>`

2.4 Adjectives, Numerals and Adverbs

CatGra m	SsCatGram	3	4	5	6	7	8	9	10	11
Adj A	Normal N	Degree	Gender	Number	Case	=	Definitene ss	TrCat		
Adj A	Cardinal C	-	-	-	Case	=	-	-		
Adj A	Ordinal O	-	-	-	Case	=	-	-		
Adv R	General G	Degree								

In the Danish tagset, the SsCatgram 'adjective' encompasses 'normal' adjectives ('AN') as well as cardinal numeral ('AC') and ordinal numerals ('AO'). Normal adjectives are marked for inflection in degree (positive/comparative/superlative/absolute superlative), gender (common/neuter), number (singular/plural), case (genitive/unmarked for case),

definiteness (definite/indefinite), and transcategorization (adverbial/unmarked for usage). Cardinal and ordinal numerals are only marked for case (genitive/unmarked for case). All normal adjectives that are not inflected in the comparative, superlative, or absolute superlative (*aller-*) forms are by default marked as being positive. Adjectives that do not inflect according to gender, number, and/or definiteness are marked with all their potential morphological values. Adjectives with 'nominal' usage (cf. Allan *et al*, 1995) can appear marked for the genitive case.

> **..en få måneder <u>gammel</u> ændring i markedsføringsloven giver flyselskaber lov til at tilbyde kunden en tilgift..**
>   <W lemma="gammel" msd="ANPCSU=IU">gammel</W>
> **..han kiggede nærmere på metalpladen og så, at det var et <u>gammelt</u> reklameskilt for Korsør Løve Margarine..**
>   <W lemma="gammel" msd="ANPNSU=IU">gammelt</W>
> **..det førte ham det meste af Fyn rundt: til <u>gamle</u> købmænd, kolonihaver, staldlofter..**
>   <W lemma="gammel" msd="ANP[CN]PU=[DI]U">gamle</W>
> **..det gav et pludseligt ryk i den <u>gamle</u> krop, og hånden hagede sig fast som en ørneklo..**
>   <W lemma="gammel" msd="ANP[CN]SU=DU">gamle</W>
> **..de er så høflige og flinke, også over for <u>ældre</u> mennesker, som ikke køber deres plader..**
>   <W lemma="gammel" msd="ANC[CN][SP]U=[DI]U">ældre</W>

When a lexical adjective is used adverbially in the PAROLE corpus, it is marked with an 'R' signifying adverbial usage, and is not tagged as an adverb.

> **..jeg anser det derfor for risikabelt at tvinge et gulv tilbage til sit <u>oprindelige</u> niveau..**
>   <W lemma="oprindelig" msd="ANP[CN]SU=DU">oprindelige</W>
> **..hvor langt dette <u>oprindelig</u> har været, er umuligt at sige, da vi ikke ved hvordan den døde har været vendt..**
>   <W lemma="oprindelig" msd="ANP- - - =- R">oprindelig</W>
> **..hvilke af aftenens sange, Nat King Cole <u>oprindeligt</u> lancerede, kan diskuteres på samme måde..**
>   <W lemma="oprindelig" msd="ANP- - - =- R">oprindeligt</W>

When distinguishing between participle, adjective and (possibly) adverb as part-of-speech tag for a given word, the distinctions made in a.o.Dansk Sprognævn, 1996 and Becker-Christensen *et al,* 1996 have generally been followed.

> **..flertallet lever stadig under plastikstykker eller tæpper, som de har <u>spændt</u> over nogle stokke..**
>   <W lemma="spænde" msd="VAPA=S[CN]I[ARU]- U">spændt</W>
> **..selvfølgelig var jeg <u>spændt</u> og nervøs for at få en kæmpegæld..**
>   <W lemma="spændt" msd="ANP[CN]SU=IU">spændt</W>
> **..lensgreve Christian Lerches meget smukke og <u>spændende</u> møntsamling..**
>   <W lemma="spændende" msd="ANP[CN][SP]U=[DI]U">spændende</W>

The Danish adverbs ('RG') are only marked for degree (positive/comparative/superlative/absolute superlative/unmarked for degree). Unlike adjectives, most adverbs are marked as 'unmarked for degree' ('U') in their base form. As mentioned above, a distinction is made between lexical adjectives used as adverbs and lexical adverbs.

..det rystede fødselsdagsbarnet fra Kaliningrad i en sådan grad, at det <u>aldrig</u> lykkedes ham at bryde tilbage..
       <W lemma="aldrig" msd="RGU">aldrig</W>

..ingen tvivl om, at ranglistens nummer 18, der <u>i søndags</u> var i finalen mod Boris Becker, ikke havde sin gode aften..
       <W lemma="i_søndags" msd="RGU">i_søndags</W>

..da hun pludselig blev ramt af en blodprop og faldt <u>bevidstløs</u> om..
       <W lemma="bevidstløs" msd="ANPCSU=IU">bevidstløs</W>

..han er nu blandt de tre danskere, der <u>frivilligt</u> er kommet hjem fra Thailand til varetægtsfængsling..
       <W lemma="frivillig" msd="ANP- - - =- R">frivilligt</W>

..Falcks svømmedykkere var hurtigt i vandet, men bilen stod <u>tom</u> på bunden af havnen..
       <W lemma="tom" msd="ANPCSU=IU">tom</W>

..hun bøjede <u>dovent</u> det ene knæ og vinkede til ham med tæerne uden at løfte ansigtet..
       <W lemma="doven" msd="ANP- - - =- R">dovent</W>

2.5 Prepositions and conjunctions

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Adpos S	**Preposition** P									
Conj C	**Coordinating** C									
Conj C	**Subordinating** S									

Both prepositions and postpositions have been tagged with 'SP' as shown the the table above. In complex prepositions, each token has been tagged separately (unlike in the Danish PAROLE lexicon). Prepositions and adverbs that form part of separable compound verbs have been tagged separately without explicit reference to their status as verbal particle.

..og <u>på</u> intet tidspunkt flyver intellektet i historien over hovedet på læseren..
       <W lemma="på" msd="SP">på</W>

..i dag har man 10 regulære forretninger, hvor der bliver solgt varer <u>af</u> præcis samme .. kvalitet..
       <W lemma="af" msd="SP">af</W>

Danish conjunctions have been classified as either coordinating conjunctions ('CC') or subordinating conjunctions ('SC').

..det er jo ikke sådan, at vi lægger op til konflikt, <u>men</u> i forvejen er udluftningen i ministeriet ikke for god..
       <W lemma="men" msd="CC">men</W>

..en sandfærdig beretning om, hvordan en arbejdsløs føler sig til grin, <u>fordi</u> han har lyst til at arbejde..
       <W lemma="fordi" msd="CS">fordi</W>

2.6 Pronouns

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Pron P	**Demonstrative** D	-	Gender	Number	Case	-	-	Register		
Pron P	**Indefinite** I	-	Gender	Number	Case	-	-	Register		
Pron P	**Interrogative/ Relative** T	-	Gender	Number	Case	-	-	Register		

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Pron P	Personal P	Person	Gender	Number	Case	-	Reflexive	Register		
Pron P	Possessive O	Person	Gender	Number	Case	Possessor	Reflexive	Register		
Pron P	Reciprocal C	-	-	Number	Case	-	-	-		

As shown in the table above, demonstrative pronouns ('PD') are marked for gender (common/neuter), number (singular/plural), case (genitive/unmarked for case), and register (obsolete/unmarked for register). As the Danish tagset does not include articles or determiners, the function of definite articles has been covered by demonstrative pronouns. *Den*, *det,* and *de* are never tagged as definite articles, but instead either as (i) demonstrative pronouns, if they appear attributively, or as (ii) personal pronouns, if they appear non- attributively. However, *denne*, *dette*, and *disse* are only tagged as demonstrative pronouns, regardless of their syntactic location. In addition, *det* is never tagged as an formal subject/object, but always as either a demonstrative or a personal pronoun.

> **..inden <u>den</u> 80- årige mand nåede til telefonen, var røveren oppe i stueetagen..**
>   <W lemma="den" msd="PD- CSU- - U">den</W>
> **..torsdag aften .. - <u>den</u> aften, de troede var deres sidste i civilisationen - ..**
>   <W lemma="den" msd="PD- CSU- - U">den</W>
> **..brug <u>den</u> i tynde flager som gratinering af en ovnret eller i tern i salatskålen..**
>   <W lemma="den" msd="PP3CSU- NU">den</W>
> **..sommer i <u>det</u> ny Europa. Ud at se med DSB - ud at lufte den indre sigøjner..**
>   <W lemma="den" msd="PD- NSU- - U">det</W>
> **..er velegnet til .. hunde og mennesker i polarkulden - og smager lige præcis af <u>det</u> smagsstof, der tilsættes..**
>   <W lemma="den" msd="PD- NSU- - U">det</W>
> **..skal man blot huske på, at man ikke må bruge madrester. <u>Det</u> vil kunne tiltrække rotter..**
>   <W lemma="det" msd="PP3NSU- NU">Det</W>

Indefinite pronouns ('PI') are marked for gender (common/neuter), number (singular/plural), case (genitive/unmarked for case) and register (obsolete/unmarked for register). The indefinite articles *en* and *et* are always tagged as indefinite pronouns in the PAROLE corpus regardless of their syntactic location. No distinction is made between *en* and *et* as pronouns or as numeral, either, instead they have always been tagged as pronouns.

> **..sætte gang i regnormenes aktivitet i havens bede. Og det er alfa og omega for <u>en</u> mere muldet jord..**
>   <W lemma="en" msd="PI- CSU- - U">en</W>
> **..i hvert fald ikke, så længe der bare er <u>én</u>, der holder af mig..**
>   <W lemma="en" msd="PI- CSU- - U">én</W>
> **..uden for teltet sidder en lille pige med <u>et</u> spædbarn på skødet..**
>   <W lemma="en" msd="PI- NSU- - U">et</W>
> **..hun stod et øjeblik stille og lyttede. Der manglede <u>et</u> eller andet..**
>   <W lemma="en" msd="PI- NSU- - U">et</W>

No distinction is made between interrogative pronouns and relative pronouns, and hence all the relevant *hv*- pronouns are tagged as interrogative/relative pronouns ('PT'). These words are marked for gender (common/neuter), number (singular/plural), case (genitive/unmarked for case), and register (unmarked for

register). In recognition of their special status in Danish, the relative pronouns *som* and *der* are tagged separately as members of the 'Unique' category.

> **..det er ligeledes interessant, <u>hvad</u> regeringen vil med redegørelsen i det udenrigspolitiske nævn..**
>> `<W lemma="hvad" msd="PT-[CN]SU--U">hvad</W>`
>
> **..Nørrebro Bibliotek introducerede for et par år siden NU-bøgerne - <u>hvilket</u> vil sige, helt aktuelle bøger..**
>> `<W lemma="hvilken" msd="PT-NSU--U">hvilket</W>`

Personal pronouns ('PP') are marked for person (1/2/3), gender (common/neuter), number (singular/plural), case (nominative/accusative/genitive), reflexiveness (yes/no), and register (polite/unmarked for register). As mentioned above, *den, det,* and *de* are always tagged as personal pronouns when they appear non-attributively, while *det* is never tagged as a formal subject/object.

> **..når jeg steg ind i <u>den</u>, var jeg stadig den pæne pige. Når jeg steg ud forvandledes jeg til vamp..**
>> `<W lemma="den" msd="PP3CSU-NU">den</W>`
>
> **..Gunnar trak <u>den</u> øverste skuffe ud og lagde brevet på bordet..**
>> `<W lemma="den" msd="PD-CSU--U">den</W>`
>
> **..nu var det alvor. Nu gjorde det ondt. Nej, det vil jeg ikke. Hvorfor <u>det</u>? - Fordi ... nåja..**
>> `<W lemma="det" msd="PP3NSU-NU">det</W>`
>
> **..Antinoos, der altid førte <u>det</u> store ord i hallen, råbte: - Hvad skal vi med alle de tiggere..**
>> `<W lemma="den" msd="PD-NSU--U">det</W>`
>
> **..derimod er <u>det</u> muligt, at firmaet, der havde videomaskinen i sin varetægt, skal erstatte den..**
>> `<W lemma="det" msd="PP3NSU-NU">det</W>`

Possessive pronouns ('PO') are marked for person (1/2/3), gender (common/neuter), number (singular/plural), case (genitive/unmarked for case), possessor (singular/plural), reflexiveness (yes/no), and register (formal/polite/unmarked for register).

> **..jeg skulle møde på <u>min</u> vagt klokken seksten, så da vi havde drukket kaffen .. måtte jeg bryde op..**
>> `<W lemma="min" msd="PO1CSUSNU">min</W>`
>
> **..Heidi sendte <u>sine</u> bedsteforældre et hurtigt blik, men så var det ligesom hun opgav..**
>> `<W lemma="sin" msd="PO3[CN]PUSYU">sine</W>`
>
> **..et værelse fyldt med legetøj og møbler, der passede i <u>hans</u> størrelse..**
>> `<W lemma="hans" msd="PO3[CN][SP]USNU">hans</W>`
>
> **..hunden gjorde i sin vildskab så voldsom modstand, at et led i <u>dens</u> halsjernlænke simpelt hen knækkede..**
>> `<W lemma="dens" msd="PO3[CN][SP]USNU">dens</W>`

Reciprocal pronouns ('PC') are marked for number (singular/plural) and case (genitive/unmarked for case). Only *hinanden* og *hinandens* have been given this tag.

..hendes arbejdsmæssige omgivelser .. kigger på <u>hinanden</u> og tænker: "Nu er hun sgu da for meget"..

      `<W lemma="hinanden" msd="PC- -PU- - -">hinanden</W>`

..søstre kender <u>hinandens</u> inderste på en måde som ingen anden kender dem..

      `<W lemma="hinanden" msd="PC- -PG- - -">hinandens</W>`

2.7 Interjections

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Interj I	=									

Interjections ('I=') are uninflected in the PAROLE corpus.

..efter gymnastikken går vi til vore forskellige sysler indtil der bli'r råbt <u>værsgo</u> !..

      `<W lemma="værsgo" msd="I=">værsgo</W>`

..<u>Aha</u> - tilbage til Bikuben. De kunne dog ikke udtale sig om sagen..

      `<W lemma="aha" msd="I=">Aha</W>`

2.8 Unique

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Unique U	=									

'Unique' ('U=') covers the infinitive marker *at,* as well as *som* and *der* to indicate their unique function. Hence, no distinction is made between *som* as a conjunction and as a relative pronoun, and between *der* as a formal subject and as a relative pronoun. Insted they have been marked as being 'unique' in all these cases.

..en lige så barsk og kynisk Harrison Ford, der også tvinges til <u>at</u> forbedre sig..

      `<W lemma="at" msd="U=">at</W>`

..en linie fra og med årgangene 1987- 88, som er ønskværdig for Chateau Belgrave <u>at</u> følge fremover..

      `<W lemma="at" msd="U=">at</W>`

..slagger fra udslukte magmakamre, <u>som</u> engang var ildovne, fulde af smeltede boblende stenmasser..

      `<W lemma="som" msd="U=">som</W>`

..det er marxismen, der er død <u>som</u> en lerdue, og gudskelov for det!..

      `<W lemma="som" msd="U=">som</W>`

..australiereren[sic] Vince Jones, <u>der</u> også kan spille lidt trompet med følsom Chet Baker- tone..

      `<W lemma="der" msd="U=">der</W>`

..desuden vil <u>der</u> være lejlighed til at se minelæggeren "Lossen", minestrygerne "Vilsund" og "Grønsund"..

      `<W lemma="der" msd="U=">der</W>`

2.9 Residual

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Residual X	Abbreviation A									
Residual X	Foreign F									
Residual X	Punctuation P									
Residual X	Formulae R									
Residual X	Symbol S									

CatGram	SsCatGram	3	4	5	6	7	8	9	10	11
Residual X	Other X									

As shown in the table above, 'Residual' covers a residual group of word which did not belong to any of the other parts-of-speech mentioned above.

Most abbreviations in the PAROLE corpus are given a proper morphosyntactic analysis, yet a few abbreviations were for various reasons tagged 'XA'.

> **..den fartglade forbryder, der fræser ud ad vort motorvejssystem med den formidable hastighed af 120 <u>km/t</u>..**
> <W lemma="km/t" msd="XA">km/t</W>
> **..86 lifter giver liftkortet til <u>FF</u> 900,- os ret til at benytte ubegrænet[sic] i en uge..**
> <W lemma="FF" msd="XA">FF</W>

As mentioned above, foreign words not included in any Danish dictionary are tagged 'XF', unless it was felt they were common enough to merit a proper morphosyntactic tag.

> **..vi var de bedste, vi var mestrene, sejren var vor, og den helt nye "Tyskland, Tyskland, <u>alles</u> <u>kaput</u>"..**
> <W lemma="alles" msd="XF">alles</W>
> <W lemma="kaput" msd="XF">kaput</W>
> **..evnen til at møde disse krav med nye ideer og en produktionskapacitet som er <u>second to none</u>..**
> <W lemma="second" msd="XF">second</W>
> <W lemma="to" msd="XF">to</W>
> <W lemma="none" msd="XF">none</W>

Punctuation marks are tagged 'XP', while various sequences of number and letters are tagged 'XR' and symbols are tagged 'XS':

> **..en atmosfære, som indeholder store mængder kuldioxid og svovlsyre<u>.</u> Himlen er orangerød..**
> <W lemma="." msd="XP">.</W>
> **..det var et menneske, som nødlandede passagerflyet <u>SK751</u> 3. juledag, ikke teknologien..**
> <W lemma="SK751" msd="XR">SK751</W>
>
> **..Venus er et meget ugæstfrit sted med en overfladetemperatur på omkring 500<u>º</u> C..**
> <W lemma="°" msd="XS">°</W>

Finally, various kinds of text errors (spelling errors, typos, inflectional errors, conjugational errors, etc.) and other kinds of non-words resulting from tokenisation errors are tagged as 'XX':

> **..køber har bestemt sig til et <u>hus/ejerlejlighed</u> til f.eks. 650.000 kr...**
> <W lemma="hus/ejerlejlighed" msd="XX">hus/ejerlejlighed</W>
> **..de mener, at Folkekongressen skal give præsidenten <u>diktaroriske</u> <u>befølelser</u>..**
> <W lemma="diktaroriske" msd="XX">diktaroriske</W> <W lemma="befølelser" msd="XX">befølelser</W>
> **..Ekstra Bladets <u>ansvarhavende</u> chefredaktør, Bent Falbert: "Det med Cosmopolitan lyder lidt mærkeligt..**
> <W lemma="ansvarhavende" msd="XX">ansvarhavende</W>

..lidt nede af den isglatte vej står et <u>folkevognsrug brød</u>, der har fragtet en håndfuld norske skibumser til St. Anton..

        <W lemma="folkevognsrug" msd="XX">folkevognsrug</W> <W lemma="brød" msd="XX">brød</W>

..det værste er at blive forfulgt <u>at</u> teenage- piger. Hvis du bare skal skrive et par autografer, er det fint

        <W lemma="at" msd="XX">at</W>

..yderligere 17 pct. af <u>udgifter</u> går til finansiering af de store teaterscener, egnsteatre og billetstøtte i København..

        <W lemma="udgifter" msd="XX">udgifter</W>

3. References

Allan, Robin, Philip Holmes and Tom Lundskær- Nielsen (1995), <u>Danish - A Comprehensive Grammar</u>, Routledge.

Becker- Christensen, Christian *et al* (1995), <u>Politikens Store Nye Nudansk Ordbog</u>, Politikens Forlag.

Dansk Sprognævn (1996), <u>Retskrivningsordbogen</u>, Aschehoug Forlag, 2. udgave.