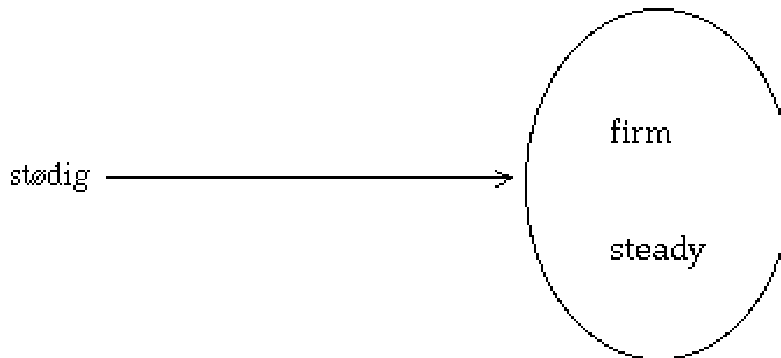


## Heuristisk utvidelse av *t*-bilder fra korpus

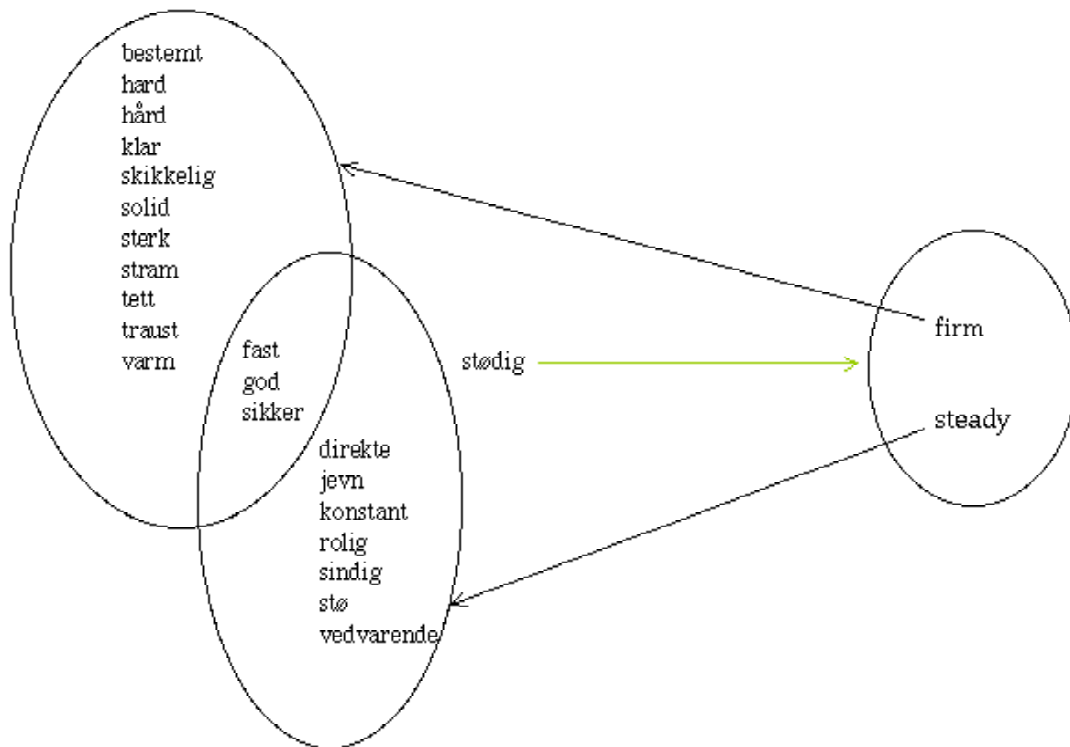
Helge Dyvik

Anta at *stødig* finnes oversatt med *firm* og *steady* i korpus:

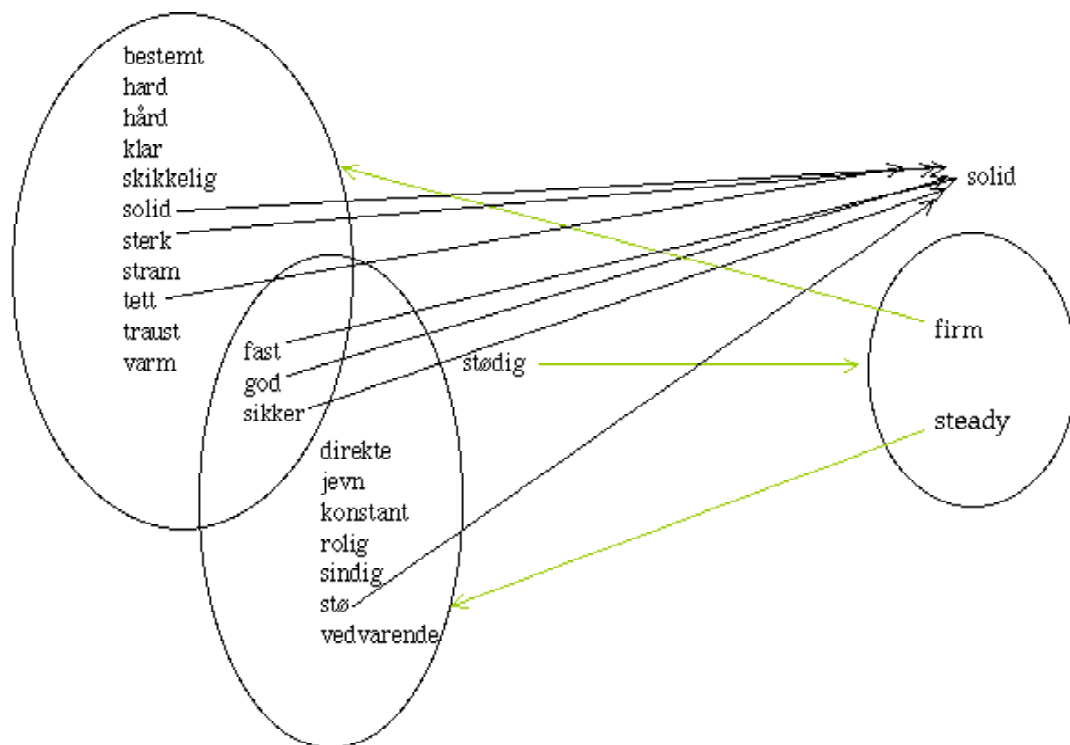


Sannsynligvis finnes det flere mulige oversettelser av *stødig*, men de er ikke med i *t*-bildet på grunn av tilfeldige hull i korpus. Grunntanken er nå at vi kan komme med gode gjetninger om slike ytterligere oversettelser ved å se på oversettelsene av *firm* og *steady* tilbake i norsk, og så se på oversettelsene av de nye norske ordene igjen. Dette må likevel gjøres med visse forbehold, som vi kommer til.

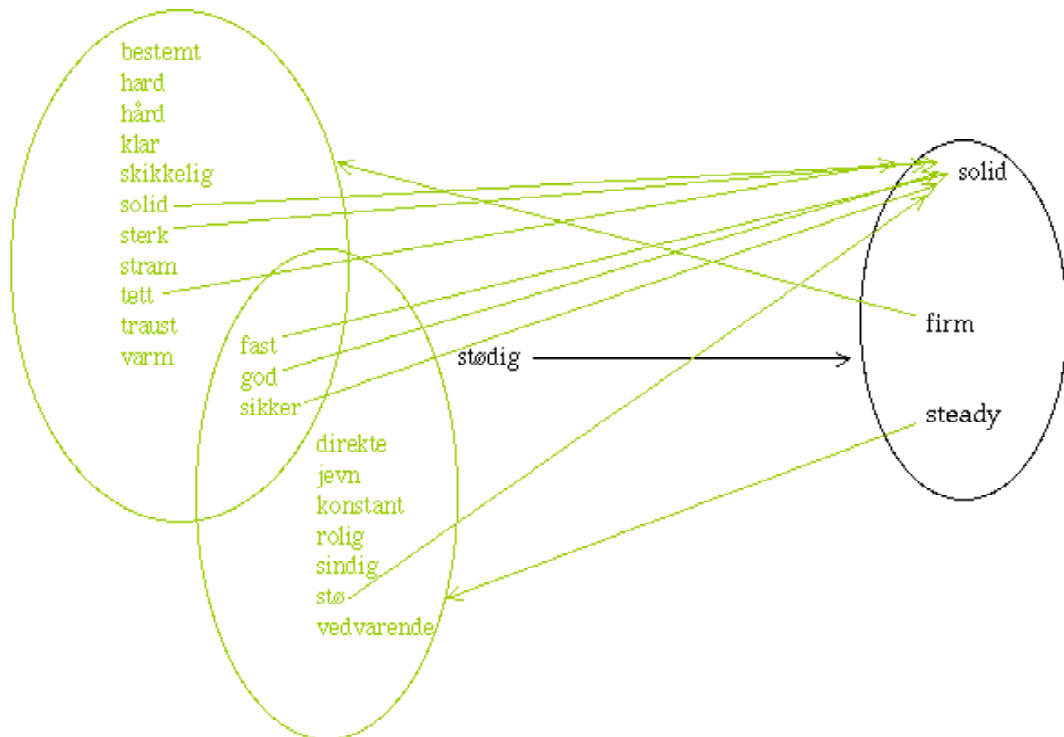
Oversettelsene av *firm* og *steady* gir oss det inverse *t*-bildet til *stødig*:



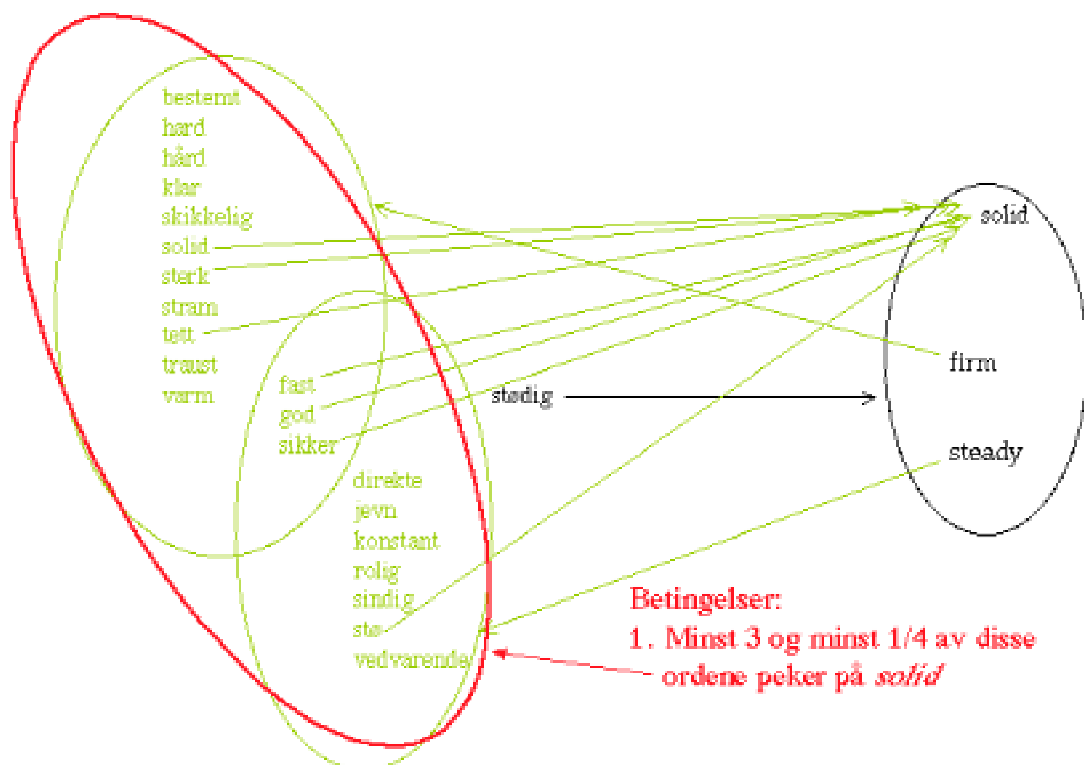
Tanken er nå: Hvis mange av ordene i det inverse bildet er «enige om» å kunne oversettes til et visst annet ord i engelsk, så er dette engelske ordet sannsynligvis også en mulig oversettelse av *stødig*. Med 'mange' kan vi f.eks. mene minst 3 og minst 1/4. Vi finner at dette gjelder det engelske ordet *solid*:



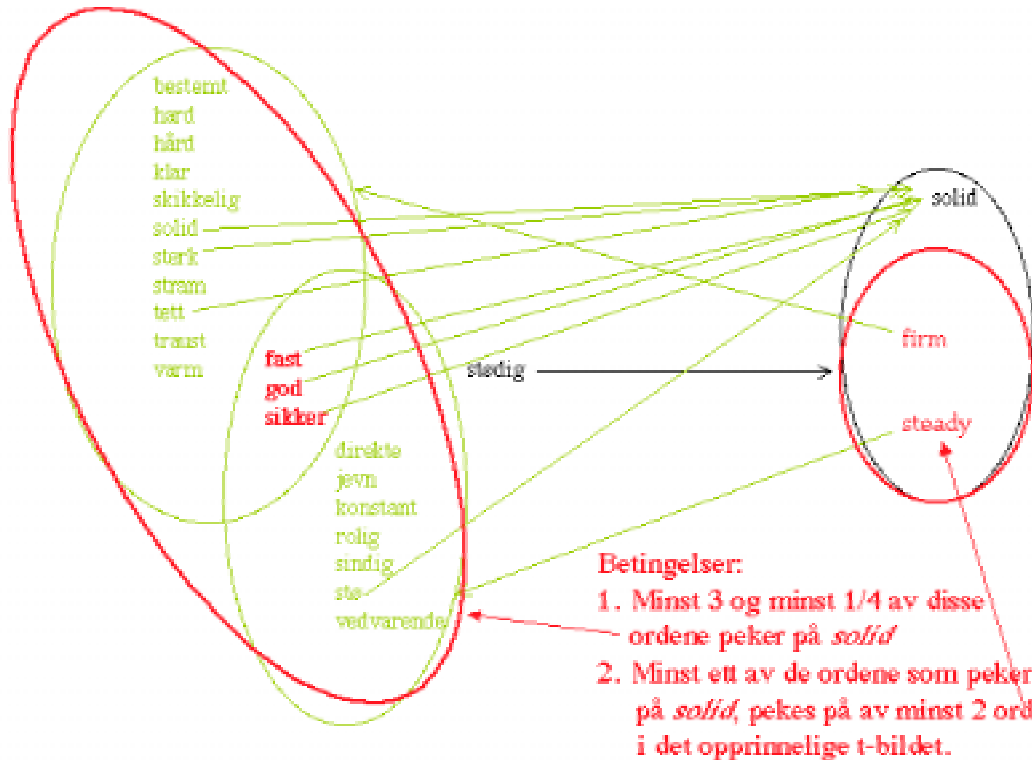
Konklusjonen er da at vi tar sjansen på å utvide det første *t*-bildet av *stødig* til også å omfatte dette nye ordet *solid*:



Betingelsene for å gjøre dette er altså for det første at minst 3 og minst 1/4 av ordene i det inverse *t*-bildet har *solid* som korrespondanse i korpus:

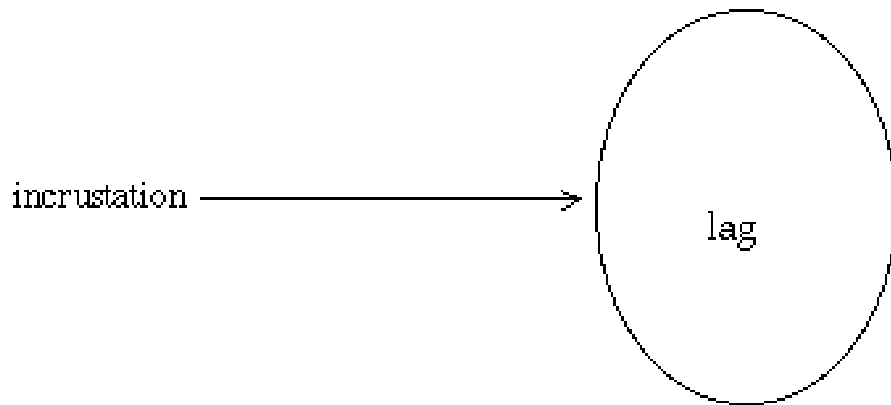


For det annet er betingelsen at minst ett av de ordene i det inverse bildet som peker på *solid*, pekes på av *minst to* ord i det første *t*-bildet. I eksempelet gjelder det hele tre ord, nemlig *fast*, *god* og *sikker*, som alle korresponderer med begge ordene i det første *t*-bildet:

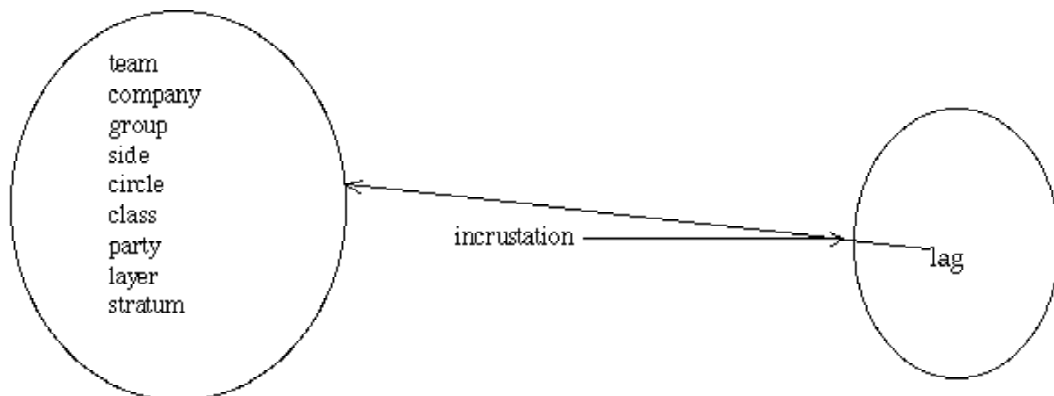


Hvorfor oppstiller vi den kompliserte betingelse nr. 2? Det er for å eliminere en feilkilde som kan ligge i *kontrastiv flertydighet* blant ordene i det primære *t*-bildet. Vi antar at kontrastiv flertydighet (altså at et ord har to eller flere innbyrdes ubeslektede betydninger – noe vi egentlig vil analysere som homonymi) er en idiosynkratisk egenskap ved enkeltord. Vi venter f.eks. ikke å finne flertydigheten mellom 'elvbredd' og 'pengeinstitusjon' i andre ord enn engelsk *bank*, hverken i engelsk eller i andre språk. Derfor er det en vesensforskjell mellom at to ord *a* og *b* er enige om å ha to felles oversettelser *u* og *v* i et annet språk, og at bare ett ord *a* har to oversettelser *u* og *v*. I sistnevnte tilfelle kan de to oversettelsene *u* og *v* være helt ubeslektet med hverandre fordi *a* er kontrastivt flertydig; i førstnevnte tilfelle er det meget lite sannsynlig, for da måtte *a* og *b* være kontrastivt flertydige på samme måte. Betingelse 2 bør dermed eliminere faren for at vi kommer frem til *solid* via en kontrastiv flertydighet hos f.eks. *firm* som ikke har noe med *stødig* å gjøre.

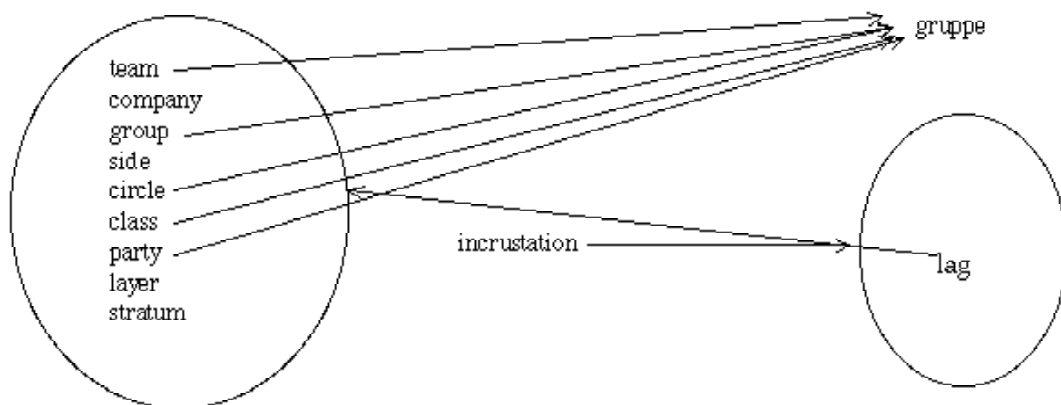
Følgende eksempel illustrerer hvordan metoden kan bære galt avgård hvis vi *ikke* forutsetter betingelse nr. 2. Engelsk *incrustation* kan oversettes til norsk *lag*:



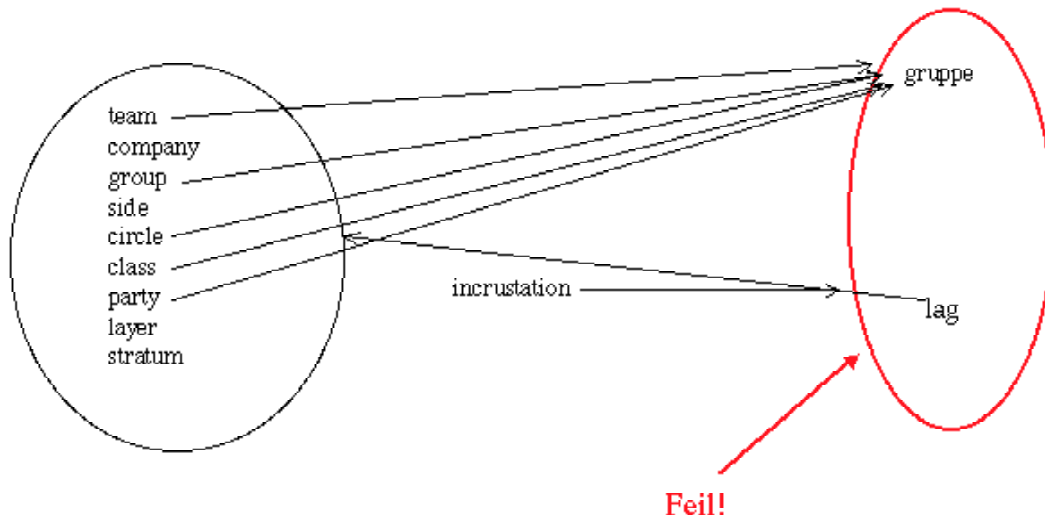
Men *lag* er kontrastivt flertydig; det kan også bety en 'gruppe', f.eks. et fotballag. Dermed får vi en mengde ord inn i det inverse *t*-bildert til *incrustation* som ikke har noe med *incrustation* å gjøre:



Mange av disse oversettes med *gruppe* i korpus:



Uten betingelse 2 ville vi nå feilaktig få *gruppe* inkludert som en mulig oversettelse av *incrustation*:



Betingelse 2 eliminerer denne feilen.

Til slutt et lite utvalg av utvidelser som resulterer av å kjøre denne rutinen på korpusmaterialet. Her er den første ordmengden etter hvert norske ord de engelske ordene man finner som oversettelser i korpus. Etter «Added:» kommer så de ordene som tentativt legges til som mulige oversettelser av den heuristiske rutinen beskrevet ovenfor. Som man ser, virker de fleste intuitivt plausible:

**fantastisk:** {amazing, brilliant, enormous, extraordinary, fantastic, glorious, great, magnificent, marvellous, remarkable, spectacular, splendid, wonderful} **Added:** {exceptional, startling, surprising, unusual}

**firma:** {business, company, establishment, firmN, place} **Added:** {factory}

**full:** {absolute, alive, bright, busy, complete, covered, crammed, crowded, drunk, full, full-time, great, grown, heavy, high, immediate, laden, loaded, packed, perfect, rich, teeming, total} **Added:** {entire, unified, whole}

**lysende:** {bright, brilliant, festive, glittering, glowing, illuminated, intense, luminous, shining} **Added:** {shiny}

**pen:** {attractive, beautiful, clean, elegant, fine, formal, gentle, good, handsome, lovely, neat, nice, pleasant, pretty, soft, sweet, tidy} **Added:** {charming, cute, dishy, enchanting, fancy, first-class, graceful, picturesque, polite}

**selskap:** {business, circle, class, companionship, company, firmN, gathering, party, society} **Added:** {cell, community, crowd, group, team}

**skikkelig:** {adequate, bad, competent, decent, effective, fair, firmA, genuine, good, nice, proper, real, regular, respectable, satisfactory, smart, sure, true} **Added:** {accurate, honest, legitimate, sincere}

**stødig:** {firm, steady} **Added:** {solid, strong}

**strålende:** {beautiful, bright, brilliant, excellent, fine, gleaming, glistening, glorious, glowing, great, magnificent, radiant, sparkling, splendid, sumptuous, warm} **Added:** {fantastic, glittering, shining, shiny}

**vedvarende:** {abiding, constant, continuous, persistent, steady, unresolved} **Added:** {perpetual}

**vennlig:** {amiable, benign, bland, cheerful, friendly, gentle, kind, kindly, nice, pleasant, pleasing, sympathetic} **Added:** {good-humoured, good-natured, polite, smiling}

**viktig:** {all-important, basic, big, central, considerable, critical, crucial, essential, great, helpful, high, important, key, large, lively, major, massive, momentous, notable, pompous, significant, urgent, vital, worthy} **Added:** {decisive, outstanding, primary, supreme}

**virkelig:** {actual, biological, confident, even, genuine, habitual, physical, proper, real, regular, right, serious, true} **Added:** {accurate, authentic, truthful}