**Algemene informatie**

Dit document bevat algemene informatie over de documenten waar jij op dit moment de beschikking over hebt. Het is niet erg als je iets niet volledig begrijpt. Probeer de vragen desondanks te beantwoorden. Als je iets echt niet begrijpt kun je de onderzoeksleider een vraag stellen.

Op een van je schermen zie je een XML bestand. Dit is de annotatie van de dialoog. Een annotatie bestaat uit drie levels.

* **Level 1.** Alle **woorden** die voorkomen in de dialoog worden gekoppeld aan woordnummers
  + w1 = hi, w2 = Jimmy, w3=hello, …
* **Level 2.** In dit level worden alle **dialoog segmenten** gespecifieerd (bestaande uit één of meerdere woorden). Dialoog segmenten zijn opgebouwd uit één of meer bovenstaande woordnummers.
* **Level 3.** In dit level wordt informatie toegevoegd aan de dialoog segmenten. Zogenaamde *dialogue acts* worden gekoppeld aan de dialoog segmenten. Een dialogue act is een combinatie die bestaat uit (**1**) een *functie* en (**2**) een *dimensie*. Verder is er mogelijk (**3**) een relatie met een of meerdere andere dialogue acts (bv. vraag-antwoord).

Een ***functie*** beschrijft de intentie die de spreker heeft met het uiten van de dialogue. Voorbeelden zijn ‘stalling’ (tijdrekken), ‘turnTake’ (de beurt pakken), en ‘answer’ (antwoord). Een functie gaat altijd samen met een dimensie.

Een ***dimensie*** beschrijft het informatie-type van de dialogue act. Zo kan een dialogue act gaan over het onderwerp van de dialoog (‘*subject’* dimensie, afkorting ‘*subj’*), over het bespreken of melden van (succesvolle of mislukte) verwerking van voorgaande uitingen (‘*feedback’* dimensie, afkorting ‘*fb’*), over het verdelen van de sprekersrol (‘*turn* *management’* dimensie, afkorting ‘*TuM’*), over het verdelen van tijd (‘*time* *management’* dimensie, afkorting ‘*TiM’*), over sociale obligaties zoals elkaar groeten (‘*social’* dimensie, afkorting ‘*soc’*), of over het corrigeren van een voorgaande uiting (‘*speech* *edit’* dimensie, afkorting ‘*spEdit’*).

Voorbeeld:

“Hallo, wat is het probleem?”

* Level 1: w1=hallo, w2=wat, w3=is, w4=het, w5=probleem.
* Level 2: dialoog segment 1 (ds1) = w1.
* Level 2: dialoog segment 2 (ds2) = w2,w3,w4,w5.
* Level 3: dialoog segment 1 (ds1) = ‘initialGreeting’ functie in ‘social’ dimensie.
* Level 3: dialoog segment 2 (ds2) = ‘openQuestion’ functie in ‘subject’ dimensie.

De volgende correcte functie-dimensie combinaties komen voor in jouw annotatie:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Functie | Dimensie |  | Functie | Dimensie |  | Functie | Dimensie |
| Inform | Subj |  | turnKeep | TuM |  | Apology | Soc |
| openQuestion | Subj |  | turnTake | TuM |  | initialGoodbye | Soc |
| answer | Subj |  | turnGrab | TuM |  | returnGoodbye | Soc |
| agreement | Subj |  | turnAssign | TuM |  | selfCorrection | spEdit |
| positiveFb | fb |  | stalling | TiM |  | correctMisspeaking | spEdit |
| checkQuestion | fb |  | initialGreeting | Soc |  |  |  |
| confirm | fb |  | returnGreeting | Soc |  |  |  |

Er kan ook sprake zijn van een **relatie** tussen dialogue acts. Een *feedback dependence relatie* is een relatie tussen een dialogue act in de feedback dimensie en een of meer andere dialogue acts. Een *functionele dependence relatie* is een relatie tussen twee of meer dialogue acts die beide geen feedback dimensie hebben (bijvoorbeeld vraag en antwoord).

**XML formaat**

De annotatie die jij gaat gebruiken is weergegeven in een XML formaat. Net als elk ander XML bestand is jouw XML bestand gestructureerd in *elementen*. Elementen bevatten een *openingstag* en een *sluitingstag* (bv. <land> Nederland</land>). Gegevens kunnen voorkomen als inhoud van een element (Nederland is inhoud van ‘naam’ element), of als *waarde* van een *attribuut*. In het voorbeeld hieronder is ‘Greenwich’ de waarde van het ‘tijdzone’ attribuut.

<datumTijd tijdzone="Greenwich">

<datum>08-sep-2005</datum>

<localeTijd>10:30 AM</localeTijd>

</datumTijd>

Jouw XML bestand is als volgt gestructureerd in drie delen die elk corresponderen met een van de annotatielevels:

<TEI>

<text>

<div>

**Level 1 - Woorden**

</div>

<div>

**Level 2 - Dialoog segmenten**

</div>

<diaml>

**Level 3 – Annotatie (met o.a. functies, dimensies en relaties)**

</diaml>

</text>

</TEI>

**Level 1.** Eerste deel van XML bestand. Eerst worden alle woorden van Anne gekoppeld aan woordnummers, daarna die van Jimmy. Dit gedeelte eindigt vlak na Jimmy’s laatste woord; w502 “Anne”. Met de opening van een nieuw <div> element begint level 2.

**Level 2.** Middelste deel van XML bestand. Hier worden alle dialoog segmenten gespecificeerd. Ze bestaan uit bovenstaande woordnummers.

<spanGrp xml:id="spGr1" type="dialogueVerbalSegment">

<span from="#w1" />

<span from="#w2" />

</spanGrp>

<fs xml:id="ds1" type="dialogueSegment">

<f name="verbalComponent" fVal="#spGrp1" />

</fs>

De woordnummers zijn te herkennen als waardes van de “from” attributen. Het “xml:id” attribuut daaronder (binnen het element “fs”) geeft aan om welk dialoog segment het gaat. Zie bovenstaand voorbeeld. Het gaat hier om dialoog segment 1 (want xml:id=“ds1”), welke bestaat uit w1 en w2 (“Hello Jimmy”).

**Level 3.** Laatste deel van XML bestand. Dit gedeelte bestaat uit ‘dialogueAct’ elementen. De dialoog segmenten zijn te herkennen als waardes van de ‘dialogSegment’ attributen. Zo zie je bijvoorbeeld dat het eerste dialogueAct element (xml:id=”da1”) gaat over dialoog segment 1 (dialogSegment=”#ds1”), dat p1 de zender (‘sender’) is en p2 de ontvanger (‘addressee’), en dat er sprake is van een ‘initialGreeting’ functie en een ‘social’ dimensie. Eventuele relaties met andere dialogue acts zijn te herkennen aan de ‘feedbackDependence’ en ‘functionalDependence’ attributen.