# SegCor: Inventory of segmentation issues concerning the TFT and Makrosyntax model

A common collection in German and French will permit to gather examples of critical segmentation cases and will help us to take decision and to try to solve these issues. In a first stage, we suppose that three examples of each type will be enough but of course feel free to add more examples or to modify critical cases titles or anything else .

|  |  |  |
| --- | --- | --- |
| Critical case | German occurrences (3 occurrences) | French occurrences (3 occurrences) |
| List format | Example 1: List expansion middle field  Transcript: Interview T191  **{**[[der rockabilly]VF [hatte]LK [sein hochkommen in den fünfzigern]MF]**V2** **|** (.) [[siebenfünfzig]KA]**KVS** **|** [[sechunfünfzig]KA]**KVS** **|** [[mit elvis presley]KA]**KVS** **|** [[den sun record studios]KA]**KVS}S**  or  **{**[[der rockabilly]VF [hatte]LK [sein hochkommen in den fünfzigern (.) siebenfünfzig sechunfünfzig mit elvis presley den sun record studios]MF]**V2}S**  Translation: *rockabilly had its uprising in the fifties, fifty-five, fifty-six, with elvis presley, the sun records*  Decision: Two possible interpretations: one complex middle field with left out connectors or a short middle field followed by verb-less chunks and expansion elements. For the maximal unit it doesn’t matter. As we follow the rule, you see is what you get, the first option will be chosen.  Example 2: Expansion with a list  Transcript: Reading to a child T14  **(1){**[[wir]VF [haben]LK [gespielt gesungen gemalt gegessen °hhh (.) und geschlafen und wieder gespielt]RK]**V2** [[plappert]LK [leo fröhlich]MF**]V1/2}C**  or  **(2){**[[wir]VF [haben]LK [gespielt]RK [gesungen gemalt gegessen °hhh (.) und geschlafen und wieder gespielt]NF]**V2** [[plappert]LK [leo fröhlich]MF**]V1/2}C**  or  **(3){**[[wir]VF [haben]LK [gespielt gesungen gemalt gegessen]RK]**V2** [[°hhh (.) und]VVF [geschlafen]RK]**V2** [[und]VVF [wieder]MF [gespielt]RK]**V2** [[plappert]LK [leo fröhlich]MF**]V1/2}C**  Translation: *we played sang painted ate and slept and played again leo babbled happily*  Decision: Interpretation whether to see it as elliptic clauses (3) or as an expanded right bracket (1) or less likely as an addition to the middle field (2) i.e. NF. | **Example 1: List of nuclei (with embedded list)**  Transcript: MJC 180-184  LEA alors ça c'est la date hein///  PAU c'est la date//  LEA c'est spa- spectacle de théâtre le vingt-quatre avril (.) avril/// prise de contact/// ce qu'il en a été:/ // les conclusions:///=  CLA =hm//  LEA euh::/ les rebondissements:/ euh:// les choix/// (0.4) l'organisation::/ euh// qui sera là/// comment on s'organise pour la: billetterie | et la buvette/ (0.5)// et puis euh:: (0.6) euh:: et puis voilà\//    Translation: *establishing contact what happened the outcome euh:: the development the choices (0.4) the organisation euh who will be there how do we get organised for ticketing and refreshments*  Decision: A series of nuclei (ungoverned elements, they do not share the presentative “c’est”), different constructions (a microsyntactic list within the last element)  **Example 2: List of nuclei (with parenthetical sequence - word search commentary)**  Transcript: Kiwi 183-186 ((03:32:00 - 03:45:95))  ELI hein normalement ben du coup moi j'ai pas  beaucoup cours/ // en fait\  [j'ai cours lundi/ ] //  BEA [ouais // finalement] //  (0.8)  ELI j'ai: interaction verbale le lun- le lundi  après-midi // (0.8) mardi j'ai pas cours //  (0.9) le mercredi j'ai euh:: (0.5)//qu'est-ce  j'ai:/ // (.) j'ai syntaxe bon // (.)  comment j'ai pu oublier ça/ //  Translation: *I have verbal interaction on Mon- on Monday afternoon (0.8) on Tuesday I have no classes (0.9) on Wednesday I have euh::: (0.5) what do I have: (.) I have syntax yeah*  Decision: A configuration of several maximal macrosyntactic units (a macro list?) |
| Oui, non, ouais…  Ja, nee | Example 1: Ja-initial  Transcript: Telephone Call T84  **(1){**[[ja ((lacht)) °h ja]KA [ich]LK [hab]LK [ihn dann so]MF [gelockt]RK]**V2}S**  or  **(2){**[[ja]KA]**KVS** [[((lacht))]KA]**KVS** [[°h ja]KA]**KVS** [[ich]LK [hab]LK [ihn dann so]MF [gelockt]RK]**V2}S**  or  **(3){**[[ja]KA]**KVS** [[((lacht))]KA]**KVS** [[°h ja]KA]**KVS}N** **{**[[ich]LK [hab]LK [ihn dann so]MF [gelockt]RK]**V2}S**  Translation: *yes ((laughs)) yes °h I then lured him this way*  Decision: (2). Yes as speech particle will be segmented separately (blue layer) from laughing and from the following material as it can theoretically stand on its own. However, on the maximum segmentation layer (red) it will be segmented together although it is analyzed like in (3). | **Example 1 :** **Turn-initial OUI**    Transcript PATAPIZZA  FLO c'est des herbes aussi//  BER oui// non mais bon//  FLO pourquoi ce racisme anti-euh:: (1.8) anti-persil///  Translation:  *FLO they are herbs too*  *BER yes no but well*  *FLO why this racism anti-euh::: (1.8) anti-parsley*  Decision: “Oui” is a nucleus, “non” may be a ligateur?  Example 2    Transcript: KIWI  ELI [.H:: ben voilà] hein// il est petit/ hein/// c'est sûr qu'il est&  MAR [trop mignon///]  ELI &plus petit que l'a:- [euh l'année pa]ssée///  BEA [ouais mais bon]//  MAR [hm/ ] //  ELI [ça va/]//  BEA [oui\ ]//    Translation:  *ELI he is small PART for sure it is&*  *MAR too cute*  *ELI &smaller that the ye- euh last year*  *BEA yeah but well*  *MAR hm*  *ELI it’s okay*  *BEA yes*  Decision: “Ouais” may be a nucleus? “oui” is a nucleus. |
| Periphery: cases other than “oui, non, ouais...” | Example 1: discourse particles (deswegen)  Transcript: Telephone Call T84  **(1){**[[jaja]KA]**KVS** [[eben]KA]**KVS** (0.22) [[nee]KA]**KVS** (.) [[deswegen]KA]**KVS** [[ich]VF [achte]LK [da eigentlich au]MF [drauf]RK]**V2}S**  or  **(2){**[[jaja]KA]**KVS** [[eben]KA]**KVS** (0.22) [[nee]KA]**KVS** (.) [[deswegen]VVF [ich]VF [achte]LK [da eigentlich au]MF [drauf]RK]**V2}S**  Translation: *yes yes exactly (0.22) no (.) therefore I’m usually careful with this*  Decision: problematic: “deswegen” is more and more used as discourse marker instead of an adverb, however still very seldomly (only about 22 cases of 1123 “deswegen” in the DGD with following V2)  Example 2: discourse marker or disruption “und dann”  Transcript: Telephone Call T84  **(1){**[[un dann]VVF (.) [die]VF [checken]LK [des ja nich]MF]**V2** [[dass]LK [es da runter]MF [geht]RK (0.42) [weißte]RAF]**VL}C**  or  **(2){**[[un]VVF [dann]VF]**KVN}A**  (.) **{**[[die]VF [checken]LK [des ja nich]MF]**V2** [[dass]LK [es da runter]MF [geht]RK (0.42) [weißte]RAF]**VL}C**  Translation: *and then (.) they don’t understand that it is deep there, you know*  Decision: ? “und dann” hasn’t been described as discourse marker yet however fulfills this function pragmatically here. | **Example 1: swearword**  Transcript: Patapizza ((02:19:53 - 02:20:68))    BER <((continue à essayer d'ouvrir le rouleau en  effeuillant son emballage, essaie de tordre  le rouleau, le regarde, pose le rouleau sur  la table, donne des coups de poings dessus,  essaie de tordre le rouleau)) (13.2)>  °putain ça pète pas\°// ((reprend le rouleau dans ses mains))    Translation:  *<((he keeps trying to open the roll in stripping his packaging, tries to twist  the roll, looks at it, lays the roll on the table, gives punches on it, tries to twist the roll)) 13.2> damn it does not break ((he picks up the roll in his hands))*  Decision: “Putain”, is it a nucleus or a prenucleus? (very common in oral corpora) |
| Disfluencies | Example 1: self correction  Transcript: Telephone Call T84  **(1){**[[ja]KA]**KVS** [[aber]VVF [das]VF [wollt]LK [ich]MF]**V2}S** **{**[[ich w]KA]**KVS}A** **{**[[äh wollt]LK [ich ja auch nie]MF]**V1/2}S**  or  **(2){**[[ja]KA]**KVS** [[aber]VVF [das]VF [wollt]LK [ich ich w]MF]**V2** [[äh wollt]LK [ich ja auch nie]MF]**V1/2}C**  Translation: *yes but that I I w uhm I never wanted that*  Decision: First option would be the strictest interpretation syntactically speaking. Second option makes more sense pragmatically. | Example 1: self correction  Transcript: MJC    BEN donc ça sera ferm- ça sera- ça sera fini à vingt-deux heures/ je pense\//    Translation:  *BEN so it will be clos- it will it will be over at ten pm I think*  Decision: Only one nucleus (the demonstrative pronoun “ça” and the auxiliary “sera” are repeated identically; the postnucleus “je pense” is shared) |
| Parenthetical  Sequence | Example 1: interjection and disruption  Transcript: Telephone Call T84  **(1){**[[also und]VVF [jetzt]VF-1]**V2-1** [[(.) oh gott]KA]**KVS** [[jet jetzt]VF-2 [weißt]LK [e]MF]**V2-2}A**  **or**  **(2){**[[also und]VVF [jetzt]VF-1]**KVN}A** **{**[[(.) oh gott]KA]**KVS}N {**[[jet jetzt]VF [weißt]LK [e]MF]**V2-2}A**  Translation: *well and now (.) oh god n now you know*  Decision: either interpretation as one (disrupted) unit with parenthesis (1) or as two disruptions with a non-sentential unit in the middle (2)  Example 1: complex discourse marker  Transcript: Telephone Call T84  **(1){**[[ich]VF [weiß]LK [au nich]MF]**V2** [[wie]LK [er das]MF [gemacht hat]RK [ehrlich gesagt]NF]**VL** [[dass]LK [er selber mit dem blumentopf]MF [runtergefallen is]RK]]**VL** [[un]VVF [der blumentopf auf ihn]MF [draufgefallen ist (.) °h]RK]]**VL}C**  **or**  **(2){**[[ich]VF [weiß]LK [au nich]MF]**V2** [[wie]LK [er das]MF [gemacht hat]RK [ehrlich gesagt]RAF]**VL** [[dass]LK [er selber mit dem blumentopf]MF [runtergefallen is]RK]]**VL** [[un]VVF [der blumentopf auf ihn]MF [draufgefallen ist (.) °h]RK]]**VL}C**  **or**  **(3){**[[ich]VF [weiß]LK [au nich]MF]**V2** [[wie]LK [er das]MF [gemacht hat]RK]**VL}C {**[[ehrlich gesagt]KA]**KVS}N {**[[dass]LK [er selber mit dem blumentopf]MF [runtergefallen is]RK]]**VL** [[un]VVF [der blumentopf auf ihn]MF [draufgefallen ist (.) °h]RK]]**VL}C**  Translation: *I don’t know how he did this - to be honest - that he fell down together with the flowerpot and the flowerpot fell onto him (cat’s paw)*  Decision: either interpretation as post field which could be integrated in the MF (1) or as a discourse marker in the right outer field of the VL-clause (2) or as independent unit which results in the segmentation of the maximal unit (3). | [see also Ex. 2 for “Macrosyntactic List Format”]  Example 1: parentheses separating a preN from the nucleus  Transcript: Kiwi 249 ((05:20:77 - 05:45:78))  ELI tu te dis °euh°/ (.) //  MAR ouais\  ELI c'est c'est compréhensif quoi\ // (0.5) et  donc elle est partie euhm: (1.3) donc elle  est partie essayer de faire ça | et tout/ //  (1.0) donc elle a cherché/ // elle a trouvé  une: formation pour ad- (0.3) adultes //  (1.3) et: euh: du coup cette formation pour  adultes // ça s'appelle comme ça donc\//  BEA ouais/ //  ELI je sai- on je sais même pas si elle est  vraiment considérée comme étudiante mais/ //  (0.6) .H donc elle s'est intégrée là-dedans  pour faire gestionnaire de paies/ (0.5)  parce qu'on avait un ami qui était là  dedans/ //  Translation: *(1.0) so she was looking for she found an ad- adult training (1.3) and: euh: in the end this adult training they call it like that [...] I kn- we I don’t even know if she is considered as a student but (0.6 )*  Decision: Two parenthetical sequences - 1st one: patently metalinguistic.  Are those two parenthetical sequences separating an ungoverned prenucleus from the rest of the maximal unit?  Or the highlighted unit is an unfinished nucleus ? |
| Reported Speech | Example 1: no verb  Transcript: Telephone Call T84  **(1){**[[ich da so]KA]**KVS** [[ja okay]KA]**KVS}N**  or  **(2){**[[ich da so]KA]**KVS}N** **{**[[ja okay]KA]**KVS}N**  or  **(3){**[[ich da so]KA [ja okay]KA]**KVS}N**  Translation: *I there so yes okay*  Decision:(3) |  |
| Boundaries in reported speech | Example 1: boundary identification  Transcript: Telephone Call T84  **(1){**[[und]VVF [dacht]LK [ich]MF]**V2** [[ja]KA]**KVS** [[gut]KA]**KVS}C {**[[des]VF [is]LK [ja scheißeg eigentlich (.) egal (.) sowieso (.) scheißegal]MF [ne]RAF]**V2}S**  **(2){**[[und]VVF [dacht]LK [ich]MF]**V2** [[ja]KA]**KVS** [[gut]KA]**KVS** [[des]VF [is]LK [ja scheißeg eigentlich (.) egal (.) sowieso (.) scheißegal]MF [ne]RAF]**V2}C**  or  **(3){**[[und]VVF [dacht]LK [ich]MF]**V2** [[ja]KA]**KVS** [[gut]KA]**KVS** [[des]VF [is]LK [ja scheißeg eigentlich (.) egal (.) sowieso (.) scheißegal]MF]**V2}C** [ne]KA]**KVS}N**  Translation: *and then I thought well yes it doesn’t matter - really doesn’t matter, right?*  Decision: The question is whether the reported speech ends with “well yes” and the rest is comment on the reported speech OR the whole belongs to the reported speech OR only the question particle in the end is commentary on the reported speech | **Example 1: graft sequence within a nucleus (layered macrosyntactic unit?) -> external boundaries**    Transcript: Kiwi 243-247    ELI c'est [bon quoi\]//  BEA [ouais/ ]//  ELI arrive un moment euh faut s` faire une idée// (0.5) et donc du coup elle s'était dit// euh: ben j` vais rentrer en formation/// pa`ce qu'elle voulait plus rentrer t` sais dans l` monde du travail| et tout [t` sais]  BEA [hm\ ]//  MAR [ouais/ ]//    Translation: *(0.5) and so in the end she said euh:: well I will start a training because she didn’t want to enter the world of work etc. anymore*  Decision:The sequence of the reported speech is followed by a governed element (the “parce que” subordinate clause). The nucleus is interrupted (noyau discontinu) ?  **Example 2: internal boundaries?**  Transcript: MJC ((06:49:05 - 07:04:87))  PAU [et là:] et là pareil euh j'ai eu un mail euh  cette semaine// elle m'a demandé où on en  était // j'ai dit que (0.5) .h ce serait pas  possible avant euh (0.4) fin mai (.) hein  éventuellement\ // (0.4) elle m'a dit //  non\// ben non // ça sera euh:: (0.5) ce sera  même pas l'année prochaine/ parce que notre  activité s'arrê[te en mars] //  SAR [ s'arrête] //    Translation: *(0.4) she told me no well no it will be euh:: (0.5) it will not even be next year because our activity stops in March*  Decision: “Non” constitutes an internal boundary for reported speech? (NOTA BENE: choral coproduction at the end, partial overlap) |
| Reported speech and disfluencies | s.o. |  |
| Complex nuclei | ? |  |
| *Segment flottant = APOKOINOU* | ? | Example 1  Transcript: MJC ((08:22:26 - 08:28:44))    LEA ben tu sais combien coûte l- le vôtre/ //  (0.3) si on avait dû vendre le:::/- le  Postscriptum/ (0.6) il coûtait deux mille  euros //    Translation: *well you know how much is y- yours (0.3) if we had to sell it the::: the Postscriptum it would cost two thousand euros*  Discussion: “Floating segment”? (right- or left-attached) |
| Subject ellipsis in coordinated sentences |  |  |
| Repairs | Example 1: interjection and self repair  Transcript: Telephone Call T84  **(1){**[[also und]VVF [jetzt]VF-1]**V2-1** [[(.) oh gott]KA]**KVS** [[jet jetzt]VF-2 [weißt]LK [e]MF]**V2-2}A**  **or**  **(2){**[[also und]VVF [jetzt]VF-1]**KVN}A** **{**[[(.) oh gott]KA]**KVS}N {**[[jet jetzt]VF [weißt]LK [e]MF]**V2-2}A**  Translation: *well and now (.) oh god n now you know*  Decision: either interpretation as one (disrupted) unit with parenthesis (1) or as two disruptions with a non-sentential unit in the middle (2) | Example 1: a new element plus previous construction  Transcript: KIWI 1282-1299  BEA oui parce que c’est un peu le hein // chez ma soeur c’est un peu la galère //  Translation: *yes because it’s quite the PART at my sister’s it’s a bit of a hassle*  Decision: The first macrosyntactic unit (N) is abandoned, the second unit present a new element and resume the previous one?  Otherwise, a whole from a macrosyntactic point of view? |
| Tag questions | Example 1:complex tag questions “weißte”  Transcript: Telephone Call T84  **(1){**[[die]VF [checken]LK [des ja nich]MF]**V2** [[dass]LK [es da runter]MF [geht]RK (0.42) [weißte]RAF]**VL}C**  or  **(2){**[[die]VF [checken]LK [des ja nich]MF]**V2** [[dass]LK [es da runter]MF [geht]RK]**VL}C** (0.42) **{**[[weißt]LK [e]MF]**V1}S**  Translation: *They don’t understand that it is deep there, you know*  Decision: ? |  |
| Parenthesis/Vocatives | Example 1: Vocative as parenthesis  Transcript: Reading to a child T14  **(1){**[[abber]VVF]**V2-1 [**[leo]KA]**KVS** (0.59) **[**[wir]VF [sin]LK [jetzt auf einem schiff]MF]**V2-2}S**  or  **(2){**[[abberleo]KA]**KVS}N** (0.59) **{[**[wir]VF [sin]LK [jetzt auf einem schiff]MF]**V2}S**  Translation: *but leo we are on a ship now*  Decision: ? |  |
| Expansions | Example 1: Expansion after speaker change  Transcript: Reading to a child T14  **(1){**[[un]VVF[wir]VF [sin]LK [einfach jetzt unterwegs (1.08) (CJ: “is toll”) mit dem schiff]MF]**V2}S**  or  **(2){**[[un]VVF[wir]VF [sin]LK [einfach jetzt unterwegs]MF]**V2}S** (1.08) (CJ: “is toll”) **{**[[mit dem schiff]KA]**KVS}N**  Translation: *and now we are just on our way (1.08) (other Speaker: “it’s great”) with the ship*  Decision: ?  Example 2: Expansion with a list  Transcript: Reading to a child T14  **(1){**[[wir]VF [haben]LK [gespielt gesungen gemalt gegessen °hhh (.) und geschlafen und wieder gespielt]RK]**V2** [[plappert]LK [leo fröhlich]MF**]V1/2}C**  or  **(2){**[[wir]VF [haben]LK [gespielt]RK [gesungen gemalt gegessen °hhh (.) und geschlafen und wieder gespielt]NF]**V2** [[plappert]LK [leo fröhlich]MF**]V1/2}C**  or  **(3){**[[wir]VF [haben]LK [gespielt gesungen gemalt gegessen]RK]**V2** [[°hhh (.) und]VVF [geschlafen]RK]**V2** [[und]VVF [wieder]MF [gespielt]RK]**V2** [[plappert]LK [leo fröhlich]MF**]V1/2}C**  Translation: *we played sang painted ate and slept and played again leo babbled happily*  Decision: Interpretation whether to see it as elliptic clauses (3) or as an expanded right bracket (1) or less likely as an addition to the middle field (2) i.e. NF. |  |
| Alternative transcription |  |  |
| Vocal communication, multimodal communication |  |  |
| Self-commenting exclamations |  |  |
| Hesitation markers and breathing |  |  |
| Collaborative turns | T14 | **Example 1: collaborative units (with particles)**  Transcript: KIWI 637-647 ((02:42:05 - 02:44:87))  BEA ben par // ah non toi on s’est croisées //  MAR dans le métro ouais //  Translation:  *BEA well from oh no you we met*  *MAR in the underground yeah*  Decision: The first macrosyntactic collaborative unit is made up of a preN (“toi”) and a nucleus (“on s’est croisées”). Peripheral elements are attached to the two collaborative units (“ah non…” and “... ouais”)? Otherwise, “ah non” is a distinct nucleus? |
|  |  |  |