

Annotation Guideline for conversational Chinese data

Table of Contents

| | |
|--|----|
| I. Introduction | 1 |
| II. A Quick Reference Section | 7 |
| III. Annotation Details | 7 |
| 3.1 Sentence boundary | 8 |
| 3.2 Linguistic phenomena to be addressed | 10 |
| 3.3 Annotation details | 12 |

I. Introduction

The conversational data consist of conversation turns between two individuals. These turns are comprised of and interrupted by incomplete fragments, fillers, disfluent sentences, discourse markers, etc. This guideline is intended to lay out the specific rules for annotating such data.

The following is a subset of a sample conversation by two speakers, marked as speaker A, and speaker B respectively:

Original:

A: “我, 我知道啊, 但是 well, 我有的时候就想, 我觉得可能因为父母为咱们受的罪就咱们应该再再就是说, 只有带带孩子的这个过程当中才能够报答父母.

B: 嗯.

A: 你说是吧? 所以我我觉得应该有孩子.

B: 那你你等以后, “

Transcribed version:

A: “我, 我知道啊, 但是 well, [noise] 我有的时候就想, 我觉得可能因为父母为咱们受的罪就咱们应该再再就是说, 只有带带孩子的这个过程当中才能够报答父母.

B: 嗯.

A: 你说是吧? 所以我我觉得应该有孩子.

B: 那你你等以后, [[distortion]]”

Note that the content with [] markings indicates non-speech events such as laughter, background noise and so on. Those in [[]] mean ...

Annotated with parts of speech and turn partitions:

((SPK 368.10~368.35~A)

(PU :)

(PN 我)

(PU ,)

(SPK 369.34~383.70~A)

(PU :)

(PN 我)

(VV 知道)

(SP 啊)

(PU ,)

(AD 但是)

(NN <English_well,>)

(NN [noise])

(PN 我)

(AD 有的时候)

(AD 就)

(VV 想)

(PU ,)

(PN 我)

(VV 觉得)

(AD 可能)

(P 因为)

(NN 父母)

(P 为)

(NN 咱们)

(VV 受)

(DEC 的)

(NN 罪)

(AD 就)

(NN 咱们)

(VV 应该)

(AD 再)

(AD 再)

(AD 就是说)

(PU ,)
(CS 只有)
(VV 带带)
(NN 孩子)
(DEC 的)
(DT 这个)
(NN 过程)
(LC 当中)
(AD 才)
(VV 能够)
(VV 报答)
(NN 父母)
(PU .)

((SPK 385.08~385.51~B)
(PU :)
(IJ 嗯)
(PU .)

((SPK 385.02~388.01~A)
(PU :)
(PN 你)
(VV 说)
(VC 是)
(SP 吧)
(PU ?)(AD 所以))
(PN 我)
(PN 我)
(VV 觉得)
(VV 应该)
(VE 有)
(NN 孩子)
(PU .))

((SPK 389.88~391.43~B)
(PU :)

(AD 那)
(PN 你)
(NN LDRB)
(PN 你)
(VV 等)
(AD 以后)
(PU ,)
(NN RDRB)
(NN [[distortion]])

LDRB and RDRB mean ... Those in [] mean ...

Fully marked and bracketed version:

((UCP (SPK 368.10~368.35~A)
 (PU :)
 (CP (DFL (NP (PN 我))))
 (PU ,)
 (SPK 369.34~383.70~A)
 (PU :)
 (IP (NP-SBJ (PN 我))
 (VP (VV 知道)))
 (SP 啊))
 (PU ,)
 (IP (ADVP (AD 但是))
 (NP-DIS (NN <English_well,>))
 (SKIP (NP (NN [noise]))))
 (IP (NP-SBJ (PN 我))
 (VP (ADVP-TMP (AD 有的时候))
 (ADVP (AD 就))
 (VP (VV 想))))
 (PU ,)
 (IP (NP-SBJ (PN 我))
 (VP (VV 觉得)
 (IP-OBJ (ADVP (AD 可能))
 (PP-PRP (P 因为)
 (NP (CP (WHNP-1 (-NONE- *OP*))

(CP (IP (NP-SBJ (NN 父母))
(VP (PP-PRP (P 为)
(NP (NN 咱们)))
(VP (VV 受)
(NP-OBJ (-NONE- *T*-1))))))
(DEC 的)))
(NP (NN 罪))))
(DFL (ADVP (AD 就)))
(INC (NP-SBJ (NN 咱们))
(VP-UNF (VV 应该)))
(DFL (ADVP (AD 再)))
(DFL (ADVP (AD 再)))
(DFL (ADVP (AD 就是说)))
(PU ,)
(NP-SBJ (-NONE- *pro*))
(VP (LCP-LOC (NP (ADVP (AD 只有))
(NP (CP (IP (NP-SBJ (-NONE- *pro*))
(VP (VV 带带)
(NP-OBJ (NN 孩子))))))
(DEC 的))
(DP (DT 这个))
(NP (NN 过程))))))
(LC 当中))
(VP (ADVP (AD 才))
(VP (VV 能够)
(VP (VV 报答)
(NP-OBJ (NN 父母)))))))))
(PU .)))

(((SPK 385.08~385.51~B)
(PU :)
(INTJ (IJ 嗯))
(PU .))

((IP (SPK 385.02~388.01~A)
 (PU :)
 (IP-Q (NP-SBJ (PN 你))
 (VP (VV 说)
 (CP-WH-OBJ (IP (NP-SBJ (-NONE- *pro*))
 (VP (VC 是)))
 (SP 吧)))
 (PU ?))
 (IP (ADVP (AD 所以))
 (DFL (NP (PN 我)))
 (NP-SBJ (PN 我))
 (VP (VV 觉得)
 (IP-OBJ (NP-SBJ (-NONE- *pro*))
 (VP (VV 应该)
 (VP (VE 有)
 (NP-OBJ (NN 孩子)))))))))
 (PU .)))

((INC (SPK 389.88~391.43~B)
 (PU :)
 (ADVP (AD 那))
 (DFL (NP (PN 你)))
 (SKIP (NP (NN LDRB)))
 (NP-SBJ (PN 你))
 (VP-UNF (VV 等)
 (ADVP-TMP (AD 以后)))
 (PU ,)
 (SKIP (NP (NN RDRB)))
 (SKIP (NP (NN [[distortion]]))))))

II. A Quick Reference Section

Phrasal labels:

FLR (i.e. filler, e.g. 嗯, 呃, 唔, 啊, 对, 这个, 那个, 就是)

DFL (disfluency followed by restarts and repair)
INC (incomplete sentences)
PRN (parenthesis; i.e. aside comments, explanations and so on)
SKIP (phrasal label; non-speech elements or transcriptions to ignore in annotation)
ETC (counterpart of ‘the like’, ‘so on and so forth’, etc. in English, also serving as final conjunct of a coordinating structure)

Functional tags:

-UNF (functional tag marking unfinished, lowest level constituents)
-DIS (i.e. discourse; functional tag attached to an adverbial phrase in contrast with other functional tags such as place, time, manner and so on),

III. Detailed discussion about annotating conversational data

This section includes three parts: sentence boundary, linguistic phenomena that need to be addressed in conversational data bracketing, and how the different phenomena are handled by specific bracketing rule and by assigning relevant phrasal labels and functional tags.

3.1 Sentence boundary

The parsing units for annotation are partitioned by sentence boundaries. A full sentence counts as a parsing unit even though this unit may contain more than one conversation turn by the same speaker, or in some occasions, by different speakers. In Example1, the turns by the same speaker are joined together as one full sentence.

Example1:

((SPK 394.17~394.34~A)
(PU :)
(NP (NN {laugh}))
(SPK 397.50~399.11~A)
(PU :)
(NP (PN 他))
(ADVP (AD 挺))

(VP (VA 好))
(SP 的)
(PU ,)
(NP (PN 他))
(ADVP (AD 就是))
(VP (VA 忙))
(PU .)
(SPK 400.09~401.03~A)
(PU :)
(ADVP (AD 就是))
(PU ,)
(SPK 401.69~403.16~A)
(PU :)
(ADVP (AD 平常))
(VP (VV 上课))
(ADVP (AD 特别))
(VP (VA 忙))
(PU ,))

In Example2, conversation turns by different speaker comprise one full parsing sentence, in which speaker A starts the conversation and speaker B interrupts and then speaker A finishes his/her sentence.

Example2:
((SPK 433.08~433.44~A)
(PU :)
(VP (VA 对))
(PU .)
(SPK 434.97~435.57~A)
(PU :)
(NP (PN 这))
(ADVP (AD 在))
(PU --))
(SPK 435.14~440.08~B)
(PU :)
(ADVP (AD 最好))
(VP (VC 是))

(VP (VE 有))
(DP (DT 那))
(CLP (M 种))
(NP (NN 项目))
(SP 的)
(SP 呀)
(PU ,)
(CC 或者是)
(NP (NN 学校))
(DEG 的)
(DP (DT 那))
(CLP (M 种))
(NP (NN 访问))
(DP (DT 那))
(CLP (M 种))
(PU ,)
(SPK 439.81~440.89~A)
(PU :)
(VP (VV 出差))
(SP 嚎)
(PU ?))

3.2 Linguistic phenomena to be addressed

3.2.1 Fillers (FLR)

Fillers have little semantic content and unrestricted distribution. They include:

嗯, 呃, 唔, 啊, 对, 这个, 那个, 就是

Example:

(1a) “然后这样以后~~就是~~, 工作呀, 然后干自己的事都方便一点”

(2a) “哦, 那你现在呢? 你还~~那个~~除了学校~~吧~~, 事情你还干别的事儿吗?”

Some of the filler words listed here, such as 嗯, 呃, 唔, 啊, can also be used as interjection words in Chinese (INTJ). They usually appear at the beginning of a sentence serving the purpose of attracting attention from the listener. When used as fillers, however, these words are meant to fill the gap, and to buy the speaker some time to complete the sentence.

3.2.2 Discourse markers (-DIS)

Discourse markers signal some relationship between the current message and the previous discourse. They bear more semantic content than fillers. The following are some examples of discourse markers:

然后 (marks a temporal succession between prior and upcoming topics in discourse; marks continuation) (e.g. ‘这边有水, 然后那边有山’)

那 (establishes the connection of and thus the relevance of the following unit of talk to a prior unit of talk) (e.g. ‘那挺好吗’)

就是说 (sometimes for further clarification, but often indicates that the speaker has got something to say)

好像 (making the utterance indirect) (eg ‘好像你得来一趟’)

反正 (holding floor, close one part of conversation and open up another part)

其实 (hedging)

我觉得

e.g.

(3a)

Speaker B: “不知道. 呃, 顺心还是挺 - 还可以 **就是说**”

Speaker A: “呃呵”

speaker B: “还可以”

3.2.3 Disfluency (DFL)

In conversations, speakers often start with something, and then repeat or abandon the content that has just been said, and then restart with revised content. This is considered disfluency. E.g.

(4a) Speaker A: ”不过他好象在这公司干干干到现在他都不想干了”

(5a) Speaker A: “他现在就是走出差这才一个月”

3.2.4 Unfinished sentences or phrases (INC vs. –UNF)

Unfinished sentences are different from fragments such as:

“九月九日 新华社 电 记者 张扬”

Examples:

(6a) “他还没有全想好就他从他现在就是走出差这才一个月”

(7a) Speaker A: “没有办法, 真是, 锋波呐? 锋”

Speaker B: “不过, 他就 –”

Speaker A: “你说”

Speaker B: “呃.”

Speaker A: “锋波现在挺好呐,”

3.2.5 Aside comments & explanations (PRN)

Aside comments often appear in natural speech. They can be viewed as content in parenthesis in writing language.

(8a) “当然是有更好的地方的话, 应该是我觉得人, 人谋活吗, 人谋活真的我觉得人应该多动一动, 应该多换一换地方,”

3.2.6 Background noises, laughter and so on (SKIP)

Laughters, background noises, lip smacks and breathing are usually marked out in the transcribed conversations.

(9a) “我, 我知道啊, 但是 well, [noise] 我有的时候就想, 我觉得可能因为父母为咱们受的罪就咱们应该再再就是说, 只有带带孩子的这个过程当中才能够报答父母。”

3.2.7 “什么/什么之类” (OTH)

(10a) “然后就那个传真号什么都有。”

3.3 Annotation details

3.3.1 Fillers (FLR)

FLR is assigned to fillers as a phrasal label. The previous utterances in Section 3.2.1 are annotated as follows:

(1b) (IP (UCP-DIS (ADVP (AD 然后))
(LCP (ADVP (AD 这样))
(LC 以后)))
(FLR (ADVP (AD 就))
(VP (VC 是)))
(PU ,) (SPK 124.21~126.97~A)
(PU :)
(IP-SBJ (NP-SBJ (-NONE- *pro*))
(VP (VP (VV 工作)
(FLR (SP 呀)))
(PU ,)
(VP (ADVP-DIS (AD 然后))
(VP (VV 干)
(NP-OBJ (DNP (NP (PN 自己))
(DEG 的))
(NP (NN 事))))))
(VP (ADVP (AD 都))
(VP (VP (VA 方便))
(ADVP (AD 一点))))

(2b) ((CP-Q (SPK 227.36~232.92~A)

(PU :)
 (CP-Q (INTJ (IJ 哦))
 (PU ,)
 (IP (ADVP-DIS (AD 那))
 (NP-SBJ (PN 你))
 (NP-TMP-PRD (NT 现在)))
 (SP 呢)
 (PU ?))
 (CP-Q (IP (DFL (NP-SBJ (PN 你))
 (ADVP (AD 还))
 (FLR (NP (PN 那个))))
 (PP-ADV (P 除了)
 (NP (NP (NN 学校))
 (FLR (SP 吧))
 (PU ,)
 (NP (NN 事情))))
 (NP-SBJ (PN 你))
 (VP (ADVP (AD 还))
 (VP (VV 干)
 (NP-OBJ (DP (DT 别的))
 (NP (NN 事儿))))))
 (SP 吗))
 (PU ?)))

3.3.2 Discourse marker (–DIS)

Discourse markers are marked out as functional tags. They serve to indicate that adverbial phrases function as discourse markers in contrast with time, location, manner, reason and so on. The examples in Section 3.2.2 are annotated as follows:

(3b) ((IP (SPK 216.96~217.62~B)
 (PU :)
 (IP (NP-SBJ (-NONE- *pro*))
 (VP (ADVP (AD 不))
 (VP (VV 知道))))
 (PU .)

(SPK 221.31~223.84~B)
 (PU :)
 (INTJ (IJ 呃))
 (PU ,)
 (IP (DFL (VP (VC 顺心))
 (ADVP (AD 还是))
 (ADVP (AD 挺))))
 (PU -)
 (NP-SBJ (-NONE- *pro*))
 (VP (ADVP (AD 还))
 (VP (VC 可以)))
 (ADVP-DIS (AD 就是说)))
 (PU ,)
 (SPK 223.92~224.49~A)
 (PU :)
 (FLR (INTJ (IJ 呃呵)))
 (PU ,)
 (SPK 224.34~225.11~B)
 (PU :)
 (IP (NP-SBJ (-NONE- *pro*))
 (VP (ADVP (AD 还))
 (VP (VC 可以))))
 (PU ,)))

3.3.3 Disfluency (DFL)

Since disfluency occurs very often, how to address this issue is quite significant in annotating such data.

A DFL unit in the annotation is skipped over. The DLF unit does NOT form a group together with its repaired/revised replacement that follows. It stands on its own, separate from the repaired element.

If the DFL unit occurs in the middle of a sentence, then it falls in wherever its repaired element is grouped. But still it doesn't form a big group with its repaired element first.

(4b) (IP (ADVP (AD 不过))
 (IP (NP-SBJ (PN 他))
 (DIS (ADVP (AD 好象)))
 (VP (PP-LOC (P 在)
 (NP (DP (DT 这))
 (NP (NN 公司)))))
 (DFL (VV 干))
 (DFL (VV 干))
 (VP (VV 干)
 (PP-TMP (P 到)
 (NP (NT 现在)))))
 (IP (NP-SBJ (PN 他))
 (VP (ADVP (AD 都))
 (ADVP (AD 不))
 (VP (VV 想)
 (VP (VV 干))))
 (SP 了)))
 (5b) (IP (NP-SBJ (PN 他))
 (VP (NP-TMP (NT 现在))
 (VP (ADVP-DIR (AD 就是))
 (DFL (VV 走))
 (VP (VV 出差)
 (NP-EXT (ADVP (AD 这))
 (ADVP (AD 才))
 (QP (CD 一)
 (CLP (M 个)))
 (NP (NN 月)))))
))))))

If the DFL unit is at the start of a sentence, then it is the first unit in an IP, CP or UCP.

(CP (DFL (NP (PN 我)))
 (PU ,)(SPK 369.34~383.70~A)
 (PU :)
 (IP (NP-SBJ (PN 我))
 (VP (VV 知道)))

(SP 啊))

More examples:

(((INC (SPK 389.88~391.43~B)
 (PU :)
 (ADVP (AD 那))
 (DFL (NP (PN 你)))
 (SKIP (NP (NN LDRB)))
 (NP-SBJ (PN 你))
 (VP-UNF (VV 等)
 (ADVP-TMP (AD 以后)))
 (PU ,)
 (SKIP (NP (NN RDRB)))
 (SKIP (NP (NN [[distortion]]))))))

(CP (IP (INTJ (IJ 呃))
 (PU ,)
 (DFL (NP (PN 我))
 (ADVP (AD 挺)))
 (PU -))
 (NP-SBJ (PN 我))
 (VP (ADVP (AD 挺))
 (VP (VV 喜欢)
 (PU +)
 (NP-PN-OBJ (NR 加拿大))))
 (PU +))
(SP 的))

(IP (NP-SBJ (PN 他))
 (VP (VV 说)
 (IP-OBJ (FLR (IJ 呃))
 (PU ,)
 (DFL (NP (NP-PN (NR 中) (NR 日)) (NP (NN 关系))))
 (FLR (SP 啊)) (PU ,)

(NP-SBJ
 (NP-PN (NR 日) (NR 中)) (NP (NN 关系)))
 (VP (NP-TMP (NT 现在)) (ADVP (AD 并不)) (ADVP (AD 那
 么))
 (PP-ADV (P 像) (PU ,) (FLR (IJ 呢)) (PU ,) (NP
 (CP (WHNP-1 (-NONE- *OP*))
 (CP
 (IP
 (NP-SBJ (NN 国际) (NN 社会))
 (VP (MSP 所)
 (VP (VV 担心) (NP-OBJ (-NONE- *T*-1))))))
 (DEC 的)))))) (VP (VA 那样)))) (PU 。)))

3.3.4 Unfinished sentences or phrases (INC. vs. –UNF)

We use INC as a top phrasal label, in contrast to top level sentence labels IP, CP or UCP, for incomplete sentences. This is different from fragments, in which cases FRAG is used for abbreviated yet interpretable sentences. INC is for sentences that's missing the last constituent, sometimes even the entire VP.

In the meantime, we use –UNF as a functional tag to mark the lowest level unfinished constituents. The combination of INC at top phrase level and –UNF as a functional tag at the lowest level can clearly show the status of the sentence and where it stops to lack the last component.

(6b) (SPK 171.20~180.71~B)
 (PU :)
 (INC (NP-SBJ (PN 他))
 (VP (ADVP (AD 还))
 (ADVP (AD 没有))
 (ADVP (AD 全))
 (VP (VV 想)
 (VP-EXT (VA 好))))))
 (ADVP (AD 就))
 (PP-UNF (P 从)))
 (IP (NP-SBJ (PN 他))

(VP (NP-TMP (NT 现在))
(VP (ADVP-DIR (AD 就是))
(DFL (VP (VV 走)))
(VP (VV 出差)
(NP-EXT (ADVP (AD 这))
(ADVP (AD 才))
(QP (CD 一)
(CLP (M 个)))
(NP (NN 月))))))

(7b) ((UCP (SPK 149.30~153.76~A)
(PU :)
(UCP (SKIP (NP (NN {laugh})))
(IP (NP-SBJ (-NONE- *pro*))
(VP (VE 没有)
(NP-OBJ (NN 办法))))
(PU ,)
(ADVP (AD 真是))
(PU ,)
(SPK 154.00~155.28~A)
(PU :)
(NP-Q (NP-PN (NR 锋波))
(SP 呐))
(PU ?)
(DFL(NP (NN 锋))))
(SPK 154.73~155.57~B)
(PU :)
(INC (ADVP (AD 不过))
(PU ,)
(NP-SBJ (PN 他))
(VP-UNF (ADVP (AD 就))))
(PU --)
(SPK 156.56~157.02~A)
(PU :)
(IP-IMP (NP-SBJ (PN 你))
(VP (VV 说)))
(PU ,)

(SPK 156.92~157.82~B)
 (PU :)
 (FLR (INTJ (IJ 呃)))
 (PU .)
 (SPK 158.30~159.45~A)
 (PU :)
 (CP-Q (IP (NP-PN-SBJ (NR 锋波))
 (VP (NP-TMP (NT 现在))
 (ADVP (AD 挺))
 (VP (VC 好))))
 (SP 呐))
 (PU ,)))

3.3.5 PRN

In the former Chinese guideline, PRN, meaning parenthesis, is already well used.

(8b) ((IP (SPK 196.31~203.16~A)
 (PU :)
 (CP-CND (IP (ADVP (AD 当然))
 (NP-SBJ (-NONE- *pro*))
 (VP (DFL (VP (VC 是)))
 (VP (VE 有)
 (NP-OBJ (DNP (ADJP (ADVP (AD 更))
 (ADJP (JJ 好)))
 (DEG 的))
 (NP (NN 地方)))))))
 (SP 的话))
 (PU ,)
 (DFL (NP-SBJ (-NONE- *pro*))
 (VP (VV 应该)
 (VP (VC 是))))
 (PRN (NP-SBJ (PN 我))
 (VP (VV 觉得)
 (IP-OBJ (IP (NP-SBJ (DFL (NP (NN 人)))
 (PU ,)

(NP (NN 人)))
 (VP (VRD (VV 谋)
 (VV 活))))
 (SP 吗)
 (PU ,)
 (IP (NP-SBJ (NN 人))
 (VP (VRD (VV 谋)
 (VV 活))))))
 (ADVP (AD 真的)))
 (NP-SBJ (PN 我))
 (VP (VV 觉得)
 (IP-OBJ (NP-SBJ (NN 人))
 (VP (VP (VV 应该)
 (VP (ADVP (AD 多))
 (VP (VV 动一动))))
 (PU ,)
 (SPK 204.28~205.61~A)
 (PU :)
 (VP (VV 应该)
 (VP (ADVP (AD 多))
 (VP (VV 换一换)
 (NP-OBJ (NN 地方))))))))
 (PU ,)))

3.3.6 SKIP

We assign SKIP to background noises, laughter, and some transcriptions that mean boundaries in the text. These are to be skipped in bracketing and annotation.

(9b) (IP (ADVP (AD 但是))
 (NP-DIS (NN <English_well,>))
 (SKIP (NP (NN [noise])))
 (IP (NP-SBJ (PN 我))
 (VP (ADVP-TMP (AD 有的时候))
 (ADVP (AD 就))
 (VP (VV 想))))

(PU ,)
 (IP (NP-SBJ (PN 我))
 (VP (VV 觉得)
 (IP-OBJ (ADVP (AD 可能))
 (PP-PRP (P 因为)
 (NP (CP (WHNP-1 (-NONE- *OP*))
 (CP (IP (NP-SBJ (NN 父母))
 (VP (PP-PRP (P 为)
 (NP (NN 咱们))))
 (VP (VV 受)
 (NP-OBJ (-NONE- *T*-1))))))
 (DEC 的)))
 (NP (NN 罪))))
 (DFL (ADVP (AD 就)))
 (INC (NP-SBJ (NN 咱们))
 (VP (VV 应该)))
 (DFL (ADVP (AD 再)))
 (DFL (ADVP (AD 再)))
 (DFL (ADVP (AD 就是说)))
 (PU ,)
 (NP-SBJ (-NONE- *pro*))
 (VP (LCP-LOC (NP (ADVP (AD 只有))
 (NP (CP (IP (NP-SBJ (-NONE- *pro*))
 (VP (VV 带带)
 (NP-OBJ (NN 孩子))))
 (DEC 的))
 (DP (DT 这个))
 (NP (NN 过程))))
 (LC 当中))
 (VP (ADVP (AD 才))
 (VP (VV 能够)
 (VP (VV 报答)
 (NP-OBJ (NN 父母)))))))))

3.3.7 OTH

In the former Chinese guideline, ETC is used as a postag for words meaning etcetera such as “等”，“等等”。But there are some other forms with similar connotations with different postags. Examples are “什么的”，“什么的”，“之类的”。These are tagged as OTH.

(10b) (IP (ADVP-DIS (AD 然后))
 (FLR (ADVP (AD 就)))
 (FLR (DP 那个))
 (NP-SBJ (NP (NN 传真号))
 (OTH (DP 什么))
 (SKIP (NP (NN LDRB))))
 (VP (ADVP (AD 都))
 (VP (VE 有))))

3.3.8 Sentential final particle (SP)

Sentence final particles are unique for Chinese language. Non-question meaning, sentence final particles are considered by some linguists as discourse markers as well. In this guideline, we only mark them as SP, not -DIS.

啊，吗/嘛，吧，呢，啦，喔，

(11) (CP-CND (IP (NP-SBJ (PN 他))
 (VP (ADVP (DFL (CS 要))
 (CS 如果))
 (VP (VE 有)
 (IP-OBJ (NP-TPC-1 (DNP (ADJP (ADVP (AD 更))
 (ADJP (JJ 好)))
 (DEG 的))
 (NP (NN 地方)))
 (PU ,)
 (NP-SBJ (PN 他))
 (VP (VV 要)
 (VP (VV 去)
 (NP-OBJ (-NONE- *T*-1)))))))))
 (SP 呐))

3.3.9 –TPC in Chinese vs. –TPC in English

The functional tag –TPC in Chinese means something different from that in English. In the English annotation guideline, -TPC means that ‘Right-dislocation is treated identically to left-dislocation; that is, the dislocated element receives the dash-tag TPC, and there is no co-indexation between the dislocated element and the argument it is associated with. Like with left-dislocation, the in situ element must be a pronoun or that. When the in situ argument is not a pronoun or that, the rightward-moved element is treated as an appositive and pseudo-attached to the in situ element with *ICH*’.

Examples are:

(12) (TOP (S (NP-SBJ (NP one)
(PP of
(NP (NP the problems)
(SBAR (WHNP-1 0)
(S (NP-SBJ they)
(VP 're
(VP facing
(NP *T*-1)
(ADVP-TMP now)
,
(NP-TPC (NP a lot)
(PP of
(NP people))))
(ADVP-TMP now))))))))))

• • • • •

(13) (TOP (S (NP-SBJ I)
 (VP guess
 (SBAR 0
 (S (NP-SBJ that)
 (VP 's
 (NP-PRD (NP a problem)
 (PP *ICH*-1))
 (ADVP too)
 (PP-1 for

(NP people))

.
(NP-TPC (NP Wait lists)
and
(NP all))))))

.
E_S))

Chinese is known as a topic-comment language, and the fronted topic in Chinese signifies what is going to be discussed in the rest of the sentence.

(14) (IP (NP-PN-TPC-1 (NR 李四))
(NP-SBJ (NR 张三))
(VP (VV 喜欢)
(NP-OBJ (-NONE- *T*-1)))
(PU .))