Annotation guidelines (Czech)

Prerequisites (general requirements):

1. create an orthographic transcription suitable for general linguistic analysis (lemmatizing, tagging, searching,...) and for the automatic generation of phonetic transcription
2. keep the number of markers/tags (symbols) at minimum (i.e. use very general ones) to reduce the necessity of interpretations made by the transcriber to get the highest possible reliability of the annotation for general use
3. chunk the recordings into chunks making coherent units containing either continuous speech, noises, breaths or silence; avoid putting the chunk boundaries into speech if possible; problems:
   a) speech overlap (there is no silence) and overlap with external noises
   b) unintelligible or ambiguous segments within otherwise intelligible (continuous) speech (see A6)
   c) speech-laughter boundaries (see B3c)
   d) speech-breath/percussive/[voc] boundaries (see e.g. B2d)
4. use separate tiers for each speaker (two speakers in the Czech recordings)
5. use third tier for noises not produced by the vocal tract of the speakers (or otherwise not clearly identifiable noises)
6. use a fourth tier for notes about extraordinary phenomena which do not fit into the annotation rules (I put there mainly question marks on sounds I do not dare to identify/classify/interpret, comments on appearance of creaky voice and sounds that I am not sure which of the speakers has produced or whether they were produced by the speakers at all, hesitations not shown otherwise or strange sounds I cannot clearly identify neither as noise, percussives nor anything else)

A Chunking:

1. chunks of speech generally correspond to continuous utterances between (in)breaths; for very long chunks (~10s) they can be divided on the silent places, but still there can be chunks of up to 6-7 sec of really continuous, uninterrupted speech
2. pauses >200ms form their own chunks
3. pauses of 100-200ms within the speech are just marked by [sil] in the transcription, but do not form their own chunks
4. in our recordings, the speakers are mixed together rather than kept on separate tracks, so that it is sometimes extremely difficult to tell who of them produced some sound (typically percussives and sometimes breaths), the preference of solutions is: a) guess and put into the most probable tier, b) put into the tiers of both speakers if really unclear; in any case add a note into the "notes" tier that explains the problem or (at least) marks the uncertainty
5. preferably create separate chunks for sounds not including speech but just groups of other sounds (breaths, percussives, filled pauses etc. / i.e. everything written in brackets) in order to keep the automatic aligner away from them - it can ignore chunks not including any speech; of course, do not create chunks for every single percussive or other non-verbal sound at the beginning or within of the speech
6. try to „cut out" problematic places (segments) within continuous sound, even if they are not

1

separated by any pause, if it can help the automatic labeller to find their position: e.g. unintelligible words or phrases within otherwise intelligible speech, intelligible words within a long laughter (where the automatic aligner would hardly be able to find them and separate from the laughter), etc.

B Basic tags:

B.1 [sil] (silent pauses)

a) used for silent pauses 100-200ms long within the speech
b) if a long chunk of speech has to be divided, the boundary will be placed behind [sil]
c) when the next word starts with a plosive, do not mark as [sil] unless the pause is really about 200ms (a normal closure of the plosive is not a pause)

B.2 [breath], [breath_in], [breath_out]

a) used for in-breaths and out-breaths (just [breath] is used where unsure)
b) includes also nasal in-/out-breaths (but see B4c!)
c) breath chunks are often very inaccurate because sometimes only the main part of the breath can be heard - if there is no clear (sharp) beginning or end (see B5d), we assume that an in-breath fills all the time (chunk) between two speech chunks; for out-breaths the boundaries have to be set rather intuitively
d) breaths usually form their own chunks, especially in-breaths between/before the speech, but out-breaths can be fluent continuations of speech when a word fluently disappears in a noisy out-breath - the last sound is long and becoming voiceless fluently (even a vowel) or - in the case of plosives - gets a kind of "aspiration" (there is usually no aspiration of plosives in Czech); in this case the [breath_out] tag has to stay in the same chunk at the end of the speech, or replaces a pause when speech immediately continues after the outhbreath (put a boundary if there is at least a short silent pause); here, subjective decision is unavoidable as to what is just a normal end of the word, what is long enough to be marked as unusual (noisy) out-breath to be included in the speech chunk (e.g. out-breath having still sound qualities (showing the formants) of the last pronounced sound) and what is really an „independent" out-breath forming a separate chunk

B.3 [laughter], [laughter_start] ... [laughter_end]

a) marks passages interpreted as laughter - this overlaps a lot with [breath] and interpretation is unavoidable: if a breath is in the context interpreted as a sign of amusement, it is a [laughter]; a laughter can end in a sharp [breath_in] just before the person starts to speak and then the in-breath can be interpreted both as a part of the laughter and an independent [breath_in] - try to separate the in-breath if possible in case of following speech
b) if at least one whole sound in a word is affected by a laughter, then [laughter_start] is placed already before this word; if the final sound is (basically) complete and just glides over into a laughter (outburst), then this is not necessary
c) if laughter continues for a longer time after the last word, try to put a boundary just after the last word (last sound), close the transcription with [laughter_end] and let the rest have it's own chunk marked as [laughter] (this breaks Prerequisite no. 3)

2

d) if there are intelligible words in a longer laughter, try to „cut them out" into their own chunks, enclosed in [laughter_start] and [laughter_end]; the rest of the laughter chunks between them are just marked as [laughter]

e) a [laughter_start] must always be followed by [laughter_end] in the same chunk (they just enclose the speech affected by laughter, do not mark the start and end of the laughter!)

f) a chunk of laughter can include the whole period of laughing, including sharp sounds, breaths (see 3a) and also longer silent pauses - if laughter continues after a silent pause (the pause is obviously a part of the laughter), the chunk will stretch across the whole period of laughter; if there seems to be just a "normal" in-breath after the pause, the laughter chunk will be closed at the beginning of the pause and the rest is just marked as an independent [breath_in] chunk (cf. B3a)

B.4 [cough], [cough_start] ... [cough_end]

a) used for any coughs, sneezes or cleaning of the throat-sounds

b) the start and end tags are used to mark words within a speech affected by e.g. cleaning the throat or nose, exactly as the [laughter...] tags

c) a (sharper) nasal [breath] (cleaning the nose?) could be also sometimes classified as a [cough]! choose [breath] unless the sound is very sharp and the function really just cleaning

B.5 [per] (percussives, unvoiced)

a) used for any percussive produced by closing and/or opening the articulators, usually using tongue or lips (clicks, lip smacks, etc.)

b) a [per] means that there is one or more(!) of such sounds within silence or other sounds, i.e. their exact position within the chunk and number is not specified; it can therefore include an unspecified amount of silence around or between them;

c) these percussives come often between an in-breath and the beginning of speech - then the click is part of the [breath_in]-chunk so that the transcription in the speech chunk can stay unaffected

d) if an in-breath starts with an audible percussive, a subjective interpretation is unavoidable: usually use "[per] [breath_in]" but if the sound is not really outstanding, take it as a natural part/start of [breath_in] and do not mark it separately; these percussives can possibly be also produced somewhere back on the palatum, velum or even glottis, but that is often difficult to classify

e) if these percussives appear within a breath, I put their tag before or after the breath tag according to where they appear relatively to the most outstanding part (peak) of the breath noise (i.e. if the loudest noise of the breath comes after the "click", then the [per] is written first and [breath] follows)

f) there can be various strange percussives in longer pauses between the speeches which do not seem to have any clear function and often not even clear origin, and so, if they are unclear or not outstanding enough, do not mark them (except of maybe putting a note on the fourth tier - conversational analysts can have a look on them later... :-)

g) marking of percussives is very unreliable in general! (sometimes it is even not clear what is a percussive of the speakers and what is some other noise)

B.6 [voc] (unarticulated voiced sounds)

a) used when the speaker produces a (voiced) unarticulated sound which cannot be classified neither as [per] nor as a clear broken word and is too short for an [ee]-filler or even recurrent (like creaky voice) – usually a false start or when the speaker prepares to speak and produces preliminary voiced sound(s) (unlike [per] which is unvoiced) that cannot be recognized as start of any word (interpretation sometimes unavoidable!)

b) this tag can also be used for sounds which are too "creaky" to build a recognizable filled pause within a speech, or any repeating unarticulated voiced sounds within a silent pause (cf. B5b)

c) a single [voc] sound can be identical with a very short filled pause like [ee] - subjective decision unavoidable! - generally it is more than just a glottal stop but less than a filled pause (and it can be interrupted)

d) a (continuous) creaky start of a sound is not marked in any way, only voiced fragments preceding an otherwise sharp start of a sound (word, speech)

B.7 [gst] (glottal stop)

a) used for a single glottal stop (without any further vowel-like sound, unlike [voc]) within a speech, which does not belong there naturally (i.e. not as a natural beginning of vowels in Czech, aso.)

b) glottal stops are neither clearly voiced nor unvoiced and could be classified either as [voc] and as [per]; the Czech automatic aligner has a trained model for a glottal stop and therefore it makes sense to transcribe a single glottal stop within a speech to give the aligner the possibility to identify it; otherwise (e.g. repeated and combined with other unarticulated sounds) it is rather classified as part of the more general [voc]-tag

C „Filled pauses" in Czech:

Three types are distinguished: [hm], [ee] and [ehe]. For pauses in speech filled by breaths use [breath] (see B2), for pauses including unarticulated voiced sounds (creaky voice or sound fragments) use [voc] (see B6), for silent pauses including percussives use [per] (see B5), for „completely" silent pauses use just [sil] (see B1).

C.1 [hm] (nasal sounds)

a) used generally for any nasal-sounding filled pause - there are different variants of nasal sounds used both as filled pauses or as back-channel responses by listeners, but often it is difficult to make a finer distinction and therefore no distinction (interpretation) is done here

C.2 [ee] (schwa sounds)

a) used for pauses filled with mainly by the „schwa" sound

b) when extremely short, can be interpreted as [voc]!

c) in a fluent speech, the filled pause can melt together with the words - if there is a (vowel-like) sound which surely doesn't fit into the text and cannot be a reduction of some other vowel or word, mark it as a filled pause

d) can end with a nasal too! as long as accent is on the schwa and not on the nasal part, it has to

4

be classified as [ee] and not as [hm]; when the filled pause is long and contains both a clear schwa and a clear nasal in comparable amounts, the combination „[ee] [hm]" is used in transcription; => can be unreliable

C.3 [ehe]

a) two-syllable sound (unlike [ee]), not used as a filler but as a back-channel signal for confirmation of acceptance of new information; equivalent of some types of [hm] pronounced just with open mouth (i.e. non-nasal pronunciation)

b) is quite clearly distinctive from the word „aha" which expresses a kind of surprise (rather than just a back-channel signal) and whose pronunciation can also be partly reduced (very seldom) but the context can help to disambiguate them quite clearly in almost all cases

D Unintelligible or ambiguous speech:

D.1 [xxx] (unintelligible speech chunk)

a) marks a segment of speech which is unintelligible, can mark a single word or many words

b) can appear in the middle of otherwise intelligible utterance (especially in speech overlap), then try to „cut it out" into it's own chunk (breaks Pre3) or - if possible - guess the most probable contents from the context and use \x-tag (see D2, following)

D.2 \x (ambiguous or hardly intelligible word)

a) tag used at the end of words that are uncertain, ambiguous or questionable in some other way (i.e. „I think the person said this, but I know I might be wrong"): if there is such a whole passage, every ambiguous word in the row is ending with this \x tag

b) used both for speech guessed plainly from a very reduced, noisy and/or unvoiced sound as well as for words/sounds/word-forms that could be also interpreted in some other way by some other transcriber or in some other context

c) if there is a chain of words marked with this tag, that implies also the fact that not only the words themselves but even the number of them can be "wrong" (=different) according to some other interpretation

d) I prefer this way of guessing over using [xxx] even for very unintelligible passages if the probable contents can be reconstructed from the context

e) this tag is also suitable for marking words and utterances where there is a mismatch (ambiguity) between the expected reaction of the speaker and the real sounds produced: e.g. when "the brain would like to 'hear' something different from what the ear can actually hear") (ref. to our discussion with Helena)

D.3 Ambiguous or unintelligible fragments of words

See H3.

E Noises:

(used on the "noise" tier only!)

E.1 [noise]

    a) a general tag for any noise that cannot be identified clearly as a sound produced by vocal tract of one of the speakers

    b) used when the noise is clearly audible and could probably make trouble for the ASR or cause distortion of the speech signal

E.2 [drawing]

    a) tag specific for our recordings where people draw pictures - a single sound made by a pencil will often stay unidentified as another "[noise]" (if clearly audible at all) but there are often long passages where the drawing can be heard as repeated noise returning in short intervals, sometimes stronger and sometimes quieter, sometimes even inaudible or with silent pauses

    b) so, this tag marks long passages where there appear repeatedly noises from the pencil (not excluding other possible noises), the length can be even 60-90 seconds, interrupted by silent pauses shorter than a few seconds (it is not possible to mark every stroke and every pause in drawing, nor decide which parts are „noisy" enough to be marked as noise or just stay unmarked as silence)

F Other problematic words:

F.1 \f (foreign word)

    a) used as tag at the end of foreign words (e.g. English) or names which are not normally used in Czech (i.e. they are here used as "quoted words" even though the pronunciation is usually adapted to Czech, unlike in German) and they keep their original spelling

F.2 \o (onomatopoeia, interjections)

    a) used as a tag at the end of onomatopoeia (usually they have no canonical orthographic form); e.g. „la\o la\o la\o", „p\o p\o pm\o", „bum\o", „tik\o tak\o"

F.3 \v (mispronunciation)

    a) used to mark words which were pronounced in a way that cannot be called reduced (not even "massively") but rather "wrong" (i.e. mispronunciation, metathesis, insertion of superfluous sounds, etc.)

F.4 \d (dialectal word form)

    a) used to mark dialectal forms that do not fit into the style or register the speaker is using otherwise (e.g. Moravian dialectal forms leaked into the speech in otherwise common Czech/Bohemian inter-dialect used by the speakers)

F.5 \p (unusual/dialectal pronunciation)

    a) used to mark words pronounced in a way that does not fit into the standard register used by

the speakers otherwise (e.g. occasional features from Moravian dialects appearing within the otherwise mostly clean bohemian)

F.6 the real pronunciation can be added to help the labeller/aligner if necessary:

a) esp. for \v in the form of [pron=...], e.g. "rohu\v[pron=horu]" means the person actually said "horu" but wanted to say "rohu" (a metathesis of syllables)
b) can be also used for foreign words (\f) (pron. adapted to the pronunciation system of the used language for the labeller), while the word itself can stay in it's original orthographic form, e.g. "correction\f[pron=korekšn]"

G Normalization of common spoken Czech:

The „common" colloquial spoken Czech is an inter-dialect used in the Czech Republic (except of the eastern provinces of Moravia and Silesia, which have their own tradition of strongly distinctive dialects and inter-dialects). It is significantly different (on all language levels) from the official norm for written language and formal communication, which is only used in speech on formal occasions or in public TV broadcasting, and less and less consequently even in private radio-stations and by teachers in the schools. The common speech can often include forms and features of both the colloquial language and the norm mixed together and there is therefore a gradual scale with a different stylistic value between the official „high standard" (which less and less people can really fully consequently follow while speaking) and the casual everyday language.

The Czech orthography can be used to write all forms of Czech and it follows rather orthoepic rules. But there is no norm for transcribing "orthographically" the common Czech spoken inter-dialect in some consistent way reducing it's own small variation and we need therefore to make our own rules according to the requirements in the „Prerequisites". The main point is to normalize the phenomena which are highly variable and even the same speakers do not follow them consequently in their own speech and often cannot be clearly distinguished in the fluent speech at all, but to keep forms which are distinctive features of the spoken language, real grammatical or lexical variants and can be always easily distinguished (if not significantly reduced, of course).

The main rule is therefore to follow the common way of writing the common spoken language (as people do more and more often when writing personal letters or messages), but normalize small variation in pronunciation in the direction of the written norm, where this can become highly variable or uncertain (i.e. normalize in the direction to a „higher" standard without loosing significant features of the real pronunciation).

G.1 Grammatical endings

Keep real grammatical endings of the spoken Czech as used in the speech (i.e. the adjective "young" in m/f/n sg. and pl. nominative: "mladej / mladá / mladý, mladý / mladý / mladý instead of the formal "mladý / mladá / mladé, mladí / mladé / mladá") - they are not pronunciation variants, but real grammatical variants and the most important feature of the spoken language.

More arguments and notes:
a) many of the endings are homonymous and the homonymy can overlap freely among completely different cases of the formal / informal language - e.g. one ending can mark different cases in the formal language and partly the same as well as completely different

cases in the informal language (see -ý or -á in the example above)

b) many common utterances and idiomatic expressions in the common speech are basically impossible to hear in the formal language so that normalization of this feature would often create an unnatural and absurd stylistic clash (e.g. to call someone stupid one can use the adjectives „hloupý" – which is neutral – or „blbý" – which belongs a lower style and which cannot occure in combination with the formal grammatical endings, so that it is normal to say (or rather write) „ty jsi hloupý", while the sentence „ty jsi blbý" would sound unnatural and funny (or dialectal?) and it would rather have the fully colloquial form „ty seš blbej")

c) the variability of formal and informal grammatical endings is systematic and the variable forms can be produced and/or identified automatically by an automatic lemmatizer if necessary (causing an additional ambiguity, of course)

G.2 -ej-

Keep diphthongization of "ý" > "ej", not only in endings but also in the root of words (i.e. "bejt, mejt, mlejt" instead of formal "být, mýt, mlít") as this is the most important and most „visible" feature of spoken Czech that everyone will notice immediately – even more than the different grammatical endings. In the mind of non-linguists, this is THE feature distinguishing formal and informal language and even many people from Moravia often think they basically speak the formal „high standard" just because they dialects lack this diphthongization.
Normalization of this feature would be also unnatural (see note G1b). Some words or expressions do not even exist or occur in the formal language (i.e. without the diphthong).

G.3 Shortening of long vowels

Normalize length of vowels in specific positions (1.p. sg. present tense of some verb classes and more...) where variation occurs (i.e. always transcribe according to the norm / formal language, e.g. "mám, prosím, podívej" instead of "mam/mám, prosim/prosím, podivej/podívej). This is a highly variable feature making a significant (finer) distinction on the stylistic scale in the common language. Long vowels are required by the norm, short vowels are used in everyday conversation. Long vowels will also be probably used consequently to raise the stylistic level e.g. when giving an informal speech, etc.
Notes and arguments:

a) the length can vary even in the common speech and sometimes it is actually hard to tell if it was pronounced as a long or as a short vowel

b) this normalization is a bit controversial step as it may look a bit "unnatural" (style) in a given context, but it is not a problem for the automatic labeller/aligner because the quality of vowels does not (mostly, except of "i/í") change with the quantity in Czech

c) this normalization also often occurs in informal written texts or transcriptions of colloquial speech

G.4 Prothetic 'v'

Delete prothetic "v" appearing in the informal pronunciation before "o" at the beginning of words / roots (i.e. write always "okno, neobvyklý" instead of "vokno/okno, nevobvyklý/neobvyklý") . This feature has probably a similar distribution and stylistic role as G3. In written form (quotations or transcriptions) it is used rather to emphasize the low stylistic level of the quoted utterance (i.e. it is probably less neutral than G3).

G.5 Initial 'j' before consonants

Transcribe initial "j" before consonants (i.e. "jsem, jméno" instead of "sem/jsem, méno/jméno").
Notes:
a) the pronunciation of this "j" is (under most circumstances) not required even in formal speech, but it is still always written and can avoid homonymy (e.g. "jsem" as a verb vs. "sem" as an adverb)
b) probably almost always written with „j" even in quotations of colloquial speech to make the text more readable, i.e. only let out in written form in order to really emphasize „authenticity" of the quotation as a spoken utterance or to emphasize the „lower style" (like G4)

G.6 Complex pronouns

Transcribe complex pronouns with the the "more official" morpheme "-hle-" instead of the wide variety of combinations "-hle- / -dle- / -le- / -hlenc- / -hlens- / -hlen- / -dlenc- / -dlens- / -dlen- / -lenc- / -lens- / -len-" (e.g. "tenhleten, touhletou, těchhletěch, takovýchhle, etc." instead of "tendleten, touletou, těchdlenctěch, takovýchlens,...") The pronunciation is often so reduced that is is even difficult to say whether there was some closure or just the fricative or even nothing at all. This wide variation could be probably classified as a combination of morphological variants and pronunciation variants, but such a classification is not known to me, the distinction would probably be difficult even between some pronunciation variants of the different morphological variants, and therefore I have avoided any distinction of these morpheme variants at all in this corpus. The number of occurrences is not so high either, so that a finer distinction could be added if some clear classification were available.

G.7 Variable verbal endings -ej(í), -aj(í)

Normalize verb forms in 3.p.pl. of the 4th class only where possible:
Some verbs have two official endings „-í/-ejí" (type „sázet") and some have only one official ending „-í" (type „prosit"). But for all those verbs there is a third (or second) alternative ending in the spoken language, which is „-ej". This ending has to be kept as „-ej" for the second type of verbs, since it is clearly distinct from „-í", but for the first type it should be normalized to „-ejí" which is probably a real pronunciation variant of „-ej", but accepted in the norm.
I.e. Distinguish both realizations „stojí / stojej" (the formal and the informal) for the latter type of verbs, but only the two official (formal) forms „nelíbějí / nelíbí" from the three possible colloquial realizations of the first type verbs. The distinction between „-ejí" and „-ej" would be often quite difficult in fluent speech anyway.
Notes:
a) This normalization also applies to other verb classes (5. class with „-ají" also can be pronounced as „-aj")...
b) There could be problem with a mixed verb type within the 4th class where many verbs newly (but not commonly) do accept the ending „-ejí" as well, i.e. they are probably on the way from the more restricted type „trpí" to the type „sází / sázejí". (see http://prirucka.ujc.cas.cz/?id=540)

G.8 Variants from other Czech dialects

Mark dialectal words with the \d tag, but normalize partial dialectal pronunciation which leaked into the speech rather because of phonetic similarity (the \p tag can be added). Alternatively add a note and/or [pron=...] tag showing the real pronunciation heard (e.g. in grammatical endings leaked from Moravian dialects "nad tou[pron=tó] poličkou[pron=poličkó]" instead of the common Czech/Bohemian (both formal and informal) „nad tou poličkou")

H Broken words, stammering and other rests of speech:

H.1 \- (word fragment or cut-off)

   a)   tag added at the end of words with a sudden cut-off or at the end of broken parts of words
   b)   for word fragments, orthoepic rules have sometimes to be followed, e.g. when the only pronounced sound is the one which is never written in the orthography (e.g. the pronunciation of "to v... vokno" could be probably written "to v\- okno" rather than "to o\- okno"; similar "já s\- jsem" rather than "já js\- jsem" for phonetic "já s...sem")
   c)   unintelligible broken fragments have to be classified as some other phenomenon: e.g. [ee], [voc], xxx, ... (see following article H3b)
   d)   the cut-off can have different forms: sharp silence, creaky voice or other [voc]-like sounds, filled-pause-like sounds, etc. => unreliable decision when to mark them additionally or just let them be represented by the cut-off symbol

H.2 stammering

   a)   if there is just one long uninterrupted sound, it is just written as if nothing special happened, but if there is an audible interruption in the sound, the stammering is transcribed as broken word fragment(s) followed by the full word (see previous article H1b for examples)

H.3 Ambiguous rests of words

   a)   if an unknown word is broken / cut off in the middle of a sound which makes it ambiguous (the sound and the word), this sound is not transcribed (e.g. it will only be transcribed "prá\-" even if the cut-off ends with a closure but it is not clear if the intended sound was a pure plosive ('t') or an affricate ('c'/'ts') or just a cut-off)
   b)   unintelligible fragments and groups of sounds have to be categorized either as [voc], [ee], [per] or just [xxx]

I   Punctuation:

The role of the punctuation is only to make the transcription more easily readable. (No syntactic rules can be followed in casual speech and following the prosody would be very difficult too.) Use only full stop and question mark.

J   "Ignored" phenomena:

   1)   „unusual" start of words or sounds – a sharp (plosive-like) start of fricatives, creaky /

fragmented start of speech (see B6d), etc.

2) "extremely long sounds" - cf. H2a

3) reduced and "massively reduced" words and utterances - written in full length

4) parts of sentences reduced to nothing or a single sound - not written at all or classified as some other phenomenon

5) creaky / whispering / unvoiced voice or speech (e.g. a person talking to himself/herself) - just mark in the "notes" tier; if (partly) intelligible, transcribed without any special tag, alternatively using \x where in doubt (see D2); if completely unintelligible, [xxx] is used (see D1) without any further marking; if really extremely silent, alternatively put a note...

6) silent and unclear noises or percussive-like sounds in longer pauses (or within a speech?) - if potentially interesting, put some note into the fourth tier (notes)

7) words interrupted "unnaturally" in the middle (e.g. extremely long closures of plosives) - ignored, as if nothing strange happened, possibly put a note into "notes"