NON-GRAMMATICAL C-UNITS IN DISCOURSE

Martin Weisser

Abstract

Most standard grammars concentrate on describing grammatically well-formed units, such as "normal" declaratives, interrogatives, etc., and only handle non-grammatical units as irregular and somehow deviant from the norm. However, in spoken language, we often encounter especially smaller or fragmentary, non-clause-like textual units that do not easily fit traditional descriptions. The aim of this article is to provide an overview of these non-grammatical units, to describe what their functions are, as well as to explain why they form such a necessary part of spoken interaction.

1   What are non-grammatical c-units?

Assuming that conventional 'sentence types' are not an appropriate means for capturing the syntactic qualities of all the elements that occur in spoken dialogues, the *Longman Grammar of Spoken and Written English* (LGr; Biber et al. 1999) operates with the concept of *c-units*. It describes these as "[…] clausal and non-clausal units […] that […] cannot be syntactically integrated with the elements that precede or follow them" (ibid.: 1070). In other words, c-units represent distinct syntactic units that, at the same time, function as independent units of sense.

Non-grammatical c-units in discourse usually correspond to what the *Comprehensive Grammar of the English Language* (CGEL; Quirk et al. 1985) treats under the heading of "Irregular Sentences" (ibid.: 89 ff.), LGr under "Unembedded dependent clauses" or "Non-clausal material" and the *Cambridge Grammar of the English Language* (CGr; Huddleston, Pullum 2002) under "Minor clause types". In general, they comprise all textual units that do not follow the rules of the standard grammatical 'sentence types', *declarative*, *interrogative*, or *imperative* in that they do not exhibit the "correct" word order or contain fewer of the elements deemed to make a sentence grammatical in the traditional sense. Some of these tend to be discussed under the heading of *exclamatives* in the traditional grammars listed above, although these are obviously not grammatically well formed in the same way as the standard sentence types.

The descriptions presented in this article are based on materials from the Spaadia trainline corpus, which was created as part of the SPAAC project (Leech & Weisser 2003) and which contains 35 speech-act annotated dialogues, altogether comprising 5,399 c-units. Although these dialogues can be classified as *task-*

oriented in that they contain operator-caller dialogues related to train timetable information and ticket bookings, many of their features are sufficiently general to allow us to generalise with regard to most other types of dialogue, too. A list of the speech act labels used in the annotation can be found in the appendix.

The importance of non-grammatical c-units in dialogue should certainly not be underestimated, since they often constitute more than 50 per cent of all c-units in a given dialogue, which can be seen from their distribution in the trainline data, where, altogether, they make up 61.53 per cent (3,322) of the units. The following graphic shows the distribution of the individual categories.
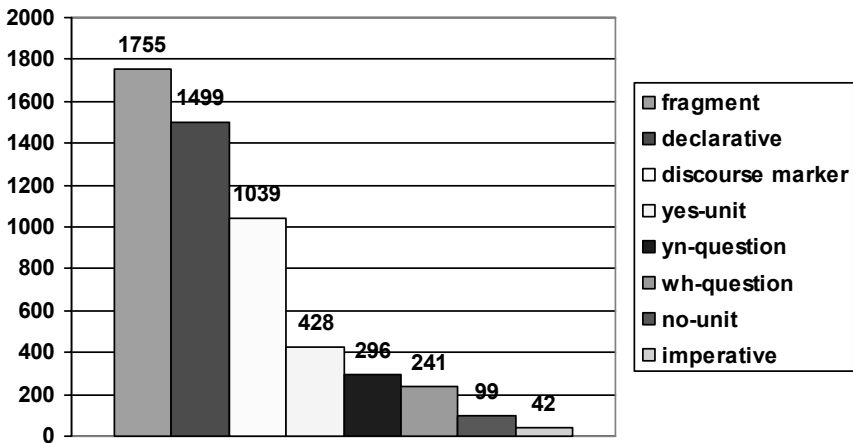

Figure 1 – Distribution of syntactic types in the trainline data

In the following sections, we will now take a closer look at the sometimes diverse functions the individual non-grammatical c-unit types can fulfil and try to explain their roles in the structuring of, and information exchange within, dialogues.

2   *Yes/no*-units

2.1 *Yes*-Units

In general, when we think about *yes* or *no* units that occur in dialogues, we usually assume that these represent answers to requests for information, just as it is generally – but equally mistakenly – assumed that questions usually contain requests for information. However, when looking at the data provided by the Spaadia corpus, it turns out that, out of 428 *yes*-units, only 22.66 per cent (97)

actually represent answers, whereas the majority of 68.46 per cent (293) simply acknowledge, 4.67 per cent (20) represent a form of acceptance, 2.8 per cent (12) constitute directives and 1.4 per cent (6) have a purely discourse marking, initiating function. Thus, although their main function undoubtedly remains to provide some kind of response to something that is expressed in some form of 'question', the data clearly show that *yes*-units are less likely to provide any real information, rather than supporting ongoing conversations. Let us take a look at some concrete examples in order to illustrate and discuss these different functions.

2.1.1 Acknowledging *yes*-units

The main function of acknowledging *yes*-units is to signal to the interlocutor that the information provided has been received or its correctness is being confirmed/reaffirmed, as in:

> <frag id="26" sp-act="inform" polarity="positive" topic="location-time-arrival" mode="deixis">arriving at 9 05 in Euston</frag></turn>
> <turn id="18" speaker="B">
> <yes id="27" sp-act="ackn">**yeah**</yes> (trainline01)

or

> <decl id="32" sp-act="confirm" polarity="positive" topic="day" mode="">
> and that's Monday the fifth</decl>
> <yes id="33" sp-act="ackn">**yeah**</yes> (trainline02)

As we can see from the two examples above, this type of acknowledgement may occur in two different ways, one where it occurs across turn boundaries, i.e. one speaker immediately acknowledges receipt of a piece of information provided by the other interlocutor, and the other a kind of reinforcement of the current speaker's own assessment of the situation within their own turn. The first type is very similar to a backchannel, the only real difference being that the acknowledging move here consists of an independent turn, rather than being "embedded" in the other speaker's turn, as it is often the case when one interlocutor only wants to signal that he or she is still following and is in agreement with what the other is saying.

2.1.2 Answering *yes*-units

Perhaps not very surprisingly, this type of *yes*-unit generally only occurs as an answer to genuine *yes/no* questions or other syntactic constructions that function as these, i.e. excluding alternative questions that simply have the syntactic shape

of the former. The following two examples provide illustrations of genuine *yes*-answers to a *yes/no*, as well as a declarative question.

```
<q-yn id="12" sp-act="reqInfo" polarity="positive" topic="number"
mode="closed">
is that two adults</q-yn></turn>
<turn id="10" speaker="B">
<yes id="13" sp-act="answ" mode="closure">yes</yes> (trainline03)
```

and

```
<decl id="44" sp-act="confirm" polarity="positive"
topic="preference-from-journey" mode="closure">you want a
ticket from Wigan now</decl>
</turn>
<turn id="20" speaker="B">
<yes id="45" sp-act="answ">yeah</yes> (trainline04)
```

If the query does contain an alternative, we need to distinguish between genuine *yes/no* questions that simply represent requests for information about a single fact, as in

```
<q-yn id="8" sp-act="reqInfo" polarity="positive" topic="creditcard"
mode="closed">
do you hold a current credit or debit card</q-yn></turn>
<turn id="6" speaker="B">
<yes id="9" sp-act="answ">yes</yes> (trainline01)
```

as opposed to alternative questions that actually ask the interlocutor to make a choice between two different options, as in the following example, where the *yes*-unit was mistakenly labelled as an answer, although it merely represents an initiating move plus elaboration:

```
<q-yn id="139" sp-act="reqInfo" polarity="positive" topic="seat"
mode="closed-alternative">is it smoking or non smoking</q-yn>
</turn>
<turn id="62" speaker="B">
<yes id="140" sp-act="answ" mode="closure">erm {#5s} yeah</yes>
<decl id="141" sp-act="answElab" polarity="negative" topic=""
mode="report">
em i'm not no actually non</decl> (trainline04)
```

2.1.3 Accepting *yes*-units

Accepting *yes*-units can basically be of two different types. On the one hand, they may be uttered in response to a directive, as in:

```
<decl id="107" sp-act="direct" polarity="positive" topic="time-departure"
mode="condition">if i ask you to be there for half an hour before departure time
of the train</decl>
</turn>
<turn id="62" speaker="B">
<yes id="108" sp-act="accept">yep</yes> (trainline14)
```

or

```
<imp id="106" sp-act="hold" polarity="positive" topic="hold"
mode="manage">hold the line</imp></turn>
<turn id="67" speaker="A">
<yes id="107" sp-act="accept">yeah</yes> (trainline35)
```

In the first example, we can see quite clearly that the directive may also be an indirect request, whereas the second example above shows a very clear directive in the form of an imperative.

On the other hand, accepting *yes*-units may be responding to offers, suggestions or requests for permission, as well as the interlocutor's providing options that may be suitable to the person responding, in which case they are similar to acknowledgments, but with an added acceptance feature, as in:

```
<decl id="40" sp-act="inform" polarity="positive" topic="number-avail"
mode="poss2">you can get the next available [ one</decl></turn>
<turn id="32" speaker="B">
<decl id="41" sp-act="appreciate" polarity="positive" topic=""
mode="appreciate">ah ] that's great</decl>
<decl id="42" sp-act="appreciate" polarity="positive" topic=""
mode="appreciate">that'd be great</decl>
<decl id="43" sp-act="accept" polarity="positive" topic="" mode="reassurance-
appreciate">that's fine</decl>
<yes id="44" sp-act="accept">yeah</yes> (trainline33).
```

This last example also illustrates that the signalling of an acceptance in form of a *yes*-unit may occasionally be difficult to recognise, unless it is accompanied by additional accepting or supporting moves in other units. In one case, however, it is almost never ambiguous, which is when the *yes* is followed by a *please* or other clearly accepting adverbial or discourse marker, such as *sure* or *ok*.

2.1.4 Directing *yes*-units

There is one exception to the above stated rule, though, which is if the unit preceding the *yes*-unit in the prior turn contains a request for a directive uttered by the previous speaker. Since this type of request always contains a query

regarding the wishes of the responding speaker, as it were placing the respondent in a "position of authority" over the person expressing the query, the *yes*-unit in this case invariably constitutes the issuing of a directive, rather than just the acceptance of a proposal, e.g.

> <q-yn id="65" sp-act="reqDirect" polarity="positive" topic="preference-booking-journey" mode="closed">do you want me go ahead and book this ticket for you</q-yn>
> </turn>
> <turn id="42" speaker="B">
> <yes id="66" sp-act="direct" mode="closure">**yes please**</yes> (trainline23)

2.1.5   Initiating/initialising *yes*-units

The remaining type of *yes*-unit is that of *yes*-units that function in a similar way to discourse markers in that they precede (initialise), and in some sense initiate, the presentation of new information. This type has the least degree of semantic content of all *yes*-units and, apart from providing a structural signal, has a more or less *phatic* and non-responding function as in:

> <dm id="41" sp-act="ackn" polarity="positive">erm right</dm>
> <yes id="42" sp-act="init" polarity="positive" topic="" mode="">**yeah**</yes>
> <q-yn id="43" sp-act="reqModal" polarity="positive" topic="time-booking" mode="closed">can i call back in 2 minutes and book that</q-yn> (trainline20)

Here, after acknowledging the previous speaker's utterance, the current speaker seems to reflect on the appropriate option, using the *yes*-unit as a kind of hesitation marker, presumably also marked by a fall-rise intonation and elongation of the word.

2.2 *No*-units

2.2.1 Answering no-units

Out of the 99 *no*-units, a much higher proportion, namely 46.46 per cent (46), constitute answers or elaborations to them, such as:

> <decl id="55" sp-act="reqInfo" polarity="negative" topic="number" mode="poss1">i can't get another one</decl></turn>
> <turn id="37" speaker="A">
> <frag id="56" sp-act="answ" polarity="negative" topic="fare" mode="">{#} not for the Virgin value fare</frag>
> <no id="57" sp-act="answElab">**no**</no> (trainline5).

In comparison to the *yes*-units, this is probably the case because *no*-units cannot have the same acknowledging or backchanneling function because they can only occur in response to negative statements, which seem to occur with much lower frequency.

2.2.2 Negating *no*-units

The next most frequent category, with 25.25 per cent (25), is that of negating *no*-units. Rather than constituting responses to queries, they correct information provided by the other interlocutor or initialise a self-correction by the current speaker, such as in:

> <decl id="8" sp-act="confirm" polarity="positive" topic="location-from-journey" mode="">and you say you're travelling from Wrexham</decl></turn>
> <turn id="6" speaker="B">
> <no id="9" sp-act="negate">**no**</no> (trainline14)

or

> <frag id="36" sp-act="answ" polarity="positive" topic="day" mode="opinion-closure">erm the next Thursday the Thursday after {#2s} which is the ninth i think {#7s}</frag>
> <no id="37" sp-act="negate">**no**</no>
> <decl id="38" sp-act="negate">it's not</decl>
> <decl id="39" sp-act="correctSelf" polarity="positive" topic="" mode="">it's the [the seventh</decl> (trainline27)

As we can see in the second example, these units are usually also followed by clarifying statements which either correct the information provided by the interlocutor or represent self-corrections.

2.2.3 Echoing *no*-units

No-units may also occur as echoes, i.e. a kind of acknowledging move to an answer, where the *no* of the previous speaker is repeated, e.g.:

> are you travelling by train today</q-yn></turn>
> <turn id="20" speaker="B">
> <no id="33" sp-act="answ" mode="closure">no</no></turn>
> <turn id="21" speaker="A">
> <no id="34" sp-act="echo">**no**</no> (trainline32)

This type of *no*-unit occurs with a frequency of 7.07 per cent (7) of all *no*-units in the corpus. An echo in general is a repetition of all or part of the information the previous speaker has uttered and may occur either as a kind of

confirming move, i.e. the information is repeated in order to verify whether it has been understood correctly, or as a type of reflection on the part of the speaker, who seems to try and absorb the information in this way and integrate it into his or her planning.

2.2.4 Directing and refusing no-units

Directing *no*-units are the negative equivalents to directing and refusing ones the counterparts to accepting *yes*-units. They both comprise 4.04 per cent (4) of the *no*-units and function essentially in the same way as their positive equivalents. In the trainline data, though, we only find refusing *no*-units that constitute reactions to offers or suggestions, such as specific available train times and none of the clear-cut, unambiguous cases that would mark these refusals, i.e. that contain the expression *no thanks/thank you*, although one *no thank you* occurs as an acknowledging move. The fact that there are no refusals to directives is probably due to the special operator-customer relationship.

2.2.5 Acknowledging *no*-units

Acknowledging *no*-units, just as their corresponding positive counterparts, occur in the two forms, within-turn and turn-spanning:

> <frag id="61" sp-act="ackn" polarity="negative" topic="date" mode="">not that date</frag>
> <no id="62" sp-act="ackn">**no**</no> (trainline04)

and

> <decl id="53" sp-act="inform" polarity="negative" topic="time-fare" mode="">the super saver is not valid at that time in the morning</decl></turn>
> <turn id="30" speaker="B">
> <no id="54" sp-act="ackn">**no**</no> (trainline16).

They make up 11.11 per cent (11) of all *no*-units and, along with their acknowledging function, always seem to signal a kind of deliberation or reflection on the part of the speaker, which is probably due to the fact that choices need to be reconsidered.

3   Discourse markers

Discourse markers (DMs), sometimes also referred to as *discourse particles*, especially if they are monomorphemic, have been extensively discussed in the literature (cf. Jucker & Ziv 1998; Fischer 2006), but there is still relatively little

agreement which words or phrases belong into this category and which do not. Often, discourse connectors, such as *therefore*, *however*, etc., are included under this heading or what I would call 'pseudo-DMs', such as *like*. The exact definition or exhaustive listing of particular discourse markers, however, is not important for our purpose here because it is possible to provide a broad framework for the description of their function without this.

In general, we can distinguish between two different types of discourse markers, those that keep a dialogue going by responding to what the other speaker has said and those that have an initiating/initialising function. In the trainline data, the latter constitute the majority of discourse markers with an overall percentage of 56.50 per cent (587) and are normally represented by the words *now*, *well*, and *so*, often preceded by a pause. Combinations of these words, also with *ok*, may occur, although *ok* on its own usually fulfils a different function, to be discussed below. Although, they do not occur in the corpus data, initial phrases like *you know* or *you see* would also belong to this group.

The other type is made up by a group of different DMs which either:

- acknowledge (369; 35.51%): *aha*, *(al)right*, *ok*,
- signal acceptance (27; 2.6%): *sure*, *ok (then)*, *right(io) (ok)*, *(ok) fine*,
- signal appreciation (10; 0.96%): *(ok) excellent*, *(ok) lovely*, *fine*,
- signal non-understanding (20; 1.92%): *pardon*, *sorry*, or
- express regret (23; 2.21%): *sorry*.

All of these either help to (re-)establish the dialogue flow or to "repair" it in some way by signalling that something may have gone wrong. The ones that belong to the former category often include the marker *ok*, and we may be able to discern a sort of cline in their strength, ranging from the basic "backchanneling" type that simply acknowledges, to the ones that signal a stronger commitment by expressing acceptance, to finally those that are more "exhuberant", which have been labelled as signalling appreciation, for want of a more explicit term. The latter category comprises expressions of apology, usually indicating that the speaker has either not understood what the preceding speaker has said or, at least in transactional dialogues, potentially regret about having provided the wrong information or made an initial erroneous choice. This kind of apologising or expressing regret is generally quite distinct from the traditional performative speech acts as originally described by Austin (1962), since they are far more interactional in nature than expressing social convention.

As we have seen for the *yes*-units discussed earlier, the occurrence of the same words in units with slightly different meanings makes it difficult to discern their exact functions easily, but at least some of the words or word combinations

may prove to be strong indicators of roughly which of the two main groups, initiating or responding, the individual markers may belong to.

4 Fragments

Fragments may arise under two different conditions: a) when it is possible semantically to only express partial information in response to queries, as part of longer information gathering processes or in conventionalised, formulaic expressions, and b) when an utterance has been abandoned before having been syntactically completed or when essential syntactic elements are unintelligible. In the former case, we usually find elliptical structures, where the missing constituents can be recovered from the context or the idiomatic character of formulaic/'ritualistic' expressions, whereas in the latter case this is impossible.

Although fragments may express nearly all types of speech acts, the majority of them (901; 51.31%) in the corpus materials is of a responding nature, comprising speech acts such as *answ* (454; 25.85%), *negate*, *echo* (230, 13.10%), *direct*, *ackn*, *accept*, *appreciate*, *correct*, *exclaim*, *expressRegret*, *pardon*, and *refer*.

The next largest group, at 337 (19.19%), consist of querying acts, such as *reqInfo* (174; 9.91%), *confirm* (161; 9.17%), and *reqDirect* (2; 0.11%), although cases of the latter are rather negligible. The predominance of the former two categories may reflect the transactional nature of the dialogue materials, though, where often smaller pieces of information are requested or repeated for confirmation and which form part of a larger "picture", e.g. all customer or travel details. Whether this may be similar in less domain-specific or non-transactional dialogue corpora remains to be verified. Examples of this are: *departing at what time* or *{#} and your initial* (reqInfo) *and from Birmingham International* or *arriving in Euston at 9 05* (confirm).

Next, we find a group of fragments (256; 14.58%) that exhibit speaker initiative in the widest sense, expressed through the speech acts *inform* (210; 11.96%), *complete*, *hold*, *init*, *expressOpinion*, *suggest*, *expressWish*, *expressPosssibility*, and *raiseIssue*.

Formulaic expressions, such as greetings and good-byes, introductions, and expressions of thanks make up 12.02 per cent (211) and, finally, completely unclassifiable ones the remaining 2.9 per cent (51). An example of the latter would be *er {#} {unclear_5_syllables} at 16 20 arriv ... arriving at 18 (33)*, where a substantial part of the information is unintelligible. One further type of fragmentary c-unit, which is usually treated in standard grammars, is that of exclamatives. However, this category does not show up in the corpus data and I can therefore not provide any descriptive statistics for this.

In terms of a general description of the syntactic structure of fragments, it is well worth noting that, very commonly, these consist of single constituents, such as NPs or PPs, of varying length and then serve as deictic references, e.g. *Monday*; *er third of October*; *er 8 minutes to 10*; *13 30*; *er {#} on {#} Monday the fifth of October at about midday*, etc. Another very common and usually slightly longer form consists of a non-finite verb plus object or PP, potentially preceded by an adverb, as in *departing at what time*; *travelling to*; *{#11s} just checking that for you {#9s}*; *arriving in Edinburgh {#} 21 08 {#}*, where, each time, the subject is omitted.

4   Conclusion

In this article, I have tried to present an overview of the different types of non-grammatical-units, together with their different functions, and, in doing so, equally to demonstrate how important they are for the functioning of naturally occurring dialogue. Knowing about them and understanding them is therefore not only a matter for scholarship or an issue in natural language processing, but also of high importance for applied linguistics and language teaching. Although my examples and statistics have been based on materials from transactional data, most of these mechanisms, apart from the exceptions already mentioned, probably also work in the same way in other types of dialogue. However, further research is still needed on data from various other domains, and especially non-transactional data, in order to understand these processes fully.

References

Austin, J. L. (1962) *How to Do Things with Words*. Oxford: Oxford University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999) *The Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

Fischer, K. (ed.) (2006) *Approaches to Discourse Particles*. Amsterdam: Elsevier.

Huddleston, R., Pullum, G. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jucker, A., Ziv, Y. (1998) 'Discourse markers: Introduction.' In: Jucker, A., Ziv, Y. (eds) *Discourse Markers: Description and Theory*. Amsterdam: John Benjamins. 1-12.

Leech, G., Weisser, M. (2003) 'Generic speech act annotation for task-oriented dialogues.' In: Archer, D., Rayson, P., Wilson, A, McEnery, T. (eds) *Proceedings of the Corpus Linguistics 2003 Conference*. Lancaster University: UCREL Technical Papers. Vol. 16. 441-446.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Appendix – List of speech acts

Speech Act Label	Brief Explanation
accept	firmly accepting
ackn	acknowledging/loosely accepting
answ	answer
answElab	elaboration to answer
appreciate	expressing appreciation; possibly accepting
bye	saying farewell; possibly closing the dialogue
complete	completing a unit begun by another party
confirm	repeating what the other party has said in order to confirm details/common ground
correct	correcting details the other party has given
correctSelf	correcting oneself
direct	giving a directive
directElab	elaboration to a directive
echo	repeating what the other party has said for purposes of verification
exclaim	expressing emotion
expressOpinion	expressing an opinion
expressPossibility	expressing possibility
expressRegret	expressing regret
expressWish	expressing a wish, i.e. potentially a mild form of directive
greet	greeting or potential uptake after a *hold*
hold	asking the other party to wait/hold the line
identifySelf	identifying oneself and/or one's institution
inform	conveying general information, or signalling awareness
informIntent	signalling the intention to do something
informIntent-hold	as above, but also asking the other party to 'hold the line'
init	initiating or initialising a new topic, sub-topic or phase in the dialogue
negate	more neutral counterpart to a refusal
offer	offering
pardon	signalling non-understanding or regret
raiseIssue	identifying an issue/a potential problem
refer	deictic reference, usually giving a time, place, etc. as an answer
refuse	refusing an offer/a proposal
reqDirect	asking for a directive
reqInfo	asking for information
reqModal	a request, which is not clearly classifiable, but contains a modal auxiliary
selfTalk	talking to oneself
suggest	making a suggestion
thank	thanking
thank-bye	thanking + saying goodbye
thirdParty	talking to an external party not directly involved in the current dialogue
unclassifiable	any speech act that does not fit any of the remaining classifications
uninterpretable	classifies a unit that is uninterpretable due to incompleteness or incoherence