Application of Psychological Characteristics to D-Script Model for Emotional Speech Processing

Artemy Kotov

Department of Theoretical Linguistics, Russian State University for Humanities, Moscow, Russia
kotov@harpia.ru
http://www.harpia.ru/d-scripts-en.html

Abstract. D-scripts model is originally developed for description of affective (emotional) mass media texts and with extension also applies to emotional speech synthesis. In this model we distinguish units for "rational" inference (r-scripts) and units for "emotional" processing of meaning (d-scripts). Basing on a psycholinguistics study we demonstrate relations between classes of emotional utterances in d-script model and psychological characteristics of informants. The study proposes a theoretical framework for an affective agent simulating given psychological characteristics in it's emotional speech behaviour.

1. Introduction

The concept of d-scripts was initially developed for the purposes of linguistic expertise and was aimed at finding emotional texts and speech insults in mass media [10]. The proposed model included two types of units (scripts) for the interpretation of incoming text: *d-scripts* (dominant scripts) for emotional processing and *r-scripts* (rational scripts) for "rational", neutral processing. It is supposed that "emotional" or "affective" texts activate d-scripts during text comprehension and tend to leave r-scripts inactive. This concept of speech analysis was further developed with orientation on CogAff architecture discussed in particular in [5, 6, 7]. CogAff (Cognition and Affect project) proposes a model for autonomous agent, experiencing and expressing emotions in interaction with the environment, so the d-scripts extension (where d-scripts correspond to "alarm system" in CogAff) offered a possibility of describe not only behavioural, but also emotional speech interaction, including speech synthesis.

CogAff architecture suggests that actual output of an agent may be controlled and affected by a hierarchy of controlling states, in particular by *personality, attitudes, preferences, moods* etc. [1]. In our study we wanted to find specific links between psychological characteristics and particular speech replies in emotional situations, defined by Rosenzweig Picture frustration test (PFT). As the d-script model provides detailed analysis of the emotional utterances, each actual reply can be associated with a specific class in d-script model, which in tern may be correlated with psychological

characteristics of a respondent, measured in parallel psychological tests. For the discovered correlations this allows to define possible classes of utterances and construct specific replies for a given set of psychological characteristics, which we want to simulate at the interface.

2. Structure and functioning of a D-script

The list of d-scripts for negative processing contains 13 units, responding for 'danger', 'limitation', 'inadequacy' and other affective meanings[1]. A particular d-script SUBJV, shown here as an example, is responsible for 'subjective actions' and is revealed in sentences (1) *You think only about yourself* - for conflict communication, and (2) *The government is concerned only about it's salary* - for influence or complaint communication [3]. Initial activation of the d-script by the listener may force utterances like (1) while an attempt to force the listener to activate the d-script may result utterances like (2). Starting model of this script fixes a situation of 'subjectivity', represented as a semantic graph (in a simple case – as a predicative structure) or in a form similar to a dictionary definition. Starting model includes slots AGGR – for person or entity, whose actions seem to be subjective, and VICT – for person, who is affected. Starting model of SUBJV is defined in the following way:

SUBJV(AGGR, VICT, M^S, P_{AGGR}, M^G): AGGR doesn't consider relevant factors of the situation and is effecting or is going to effect [all the possible] actions P_{AGGR} upon discovering of the stimulus M^S or to achieve a goal M^G; AGGR and VICT are linked with a relation $R_{AGGR-VICT}$.

Some nodes of the representation are also appended by semantic markers, responsible for emotional processing – *critical elements*. The value of the critical elements allows to distinguish emotional text and a neutral text containing the same semantic graph. The analysis of mass media and conflict texts gives us a notion on the list of critical elements for each d-script. Above all, the following critical elements are relevant to SUBJV: **<number of AGGR>**$^+$ (*Everybody thinks only about himself!*), **<timeframe/quantity of P_{AGGR}>**$^+$ (*He always talks about football!*), **<intensity of P_{AGGR}>**$^+$ (*Why do you start shouting, when I mention the washing machine?*), **<importance of M_2>**$^-$ (*Luzhkov shall bite to death everybody in order to be the first to congratulate Eltsin with his birthday!*). Definition of each d-script includes from 6 to 20 critical elements, discovered from actual text analysis and collected from linguistic studies of emotional texts [8, 9]. Critical elements may apply during text synthesis in emotional state and further may be extracted from text and contribute to emotional processing during text comprehension. In linguistic analysis meaning shifts in critical elements serve as criteria to identify emotional texts.

The sensitivity of a d-script may vary, depending on the control states: in nervous condition the of d-scripts are pre-activated and may respond to texts having very little shifts in critical elements or no shifts at all (neutral texts). In reply, the system may respond with a more emotional utterance, supporting a dialogue as:

(3) – *The government is working on the budget*.

[1] For the list of negative d-scripts see: http://www.harpia.ru/d-scripts-en.html

– They all always shout only about their budget/such trifles!

Activation of a d-script may result text synthesis not only from the starting model (description of "a terrible situation") but also from target models of aggression or flight (*I shall kill you! / I have to go away!*). Out purpose is to test the dependency between such control states as personal characteristics and the preferency of d/r-scripts in a psychological survey study.

3. Application of d-script model to the results of Rosenzweig PFT

In standard Rosenzweig PFT a participant is asked to reply to a speaker in situation of frustration, represented by a picture (the test is projective). Sample situations include:

a. [a woman tells a person, who broke a vase:] *You broke the favourite vase of my mother!* ('vase' situation) and

b. [a driver tells a pedestrian:] *I'm sorry to spoil your suite, although I tried to avoid the puddle!* ('driver' situation).

In a task of incoming text processing a d-script must detect a predicative structure, so application of the model to the processing of a situation (as in PFT) needs some extensions as no predicative structure is given a priori. In this case activation of d/r-scripts is based on a segmentation of the situation with further extraction of a typical predicative structure from each of the segments. The analysis of the set of utterances received in the experiment shows that the following segmentation of a situation appears sufficient to describe most of the actual utterances.

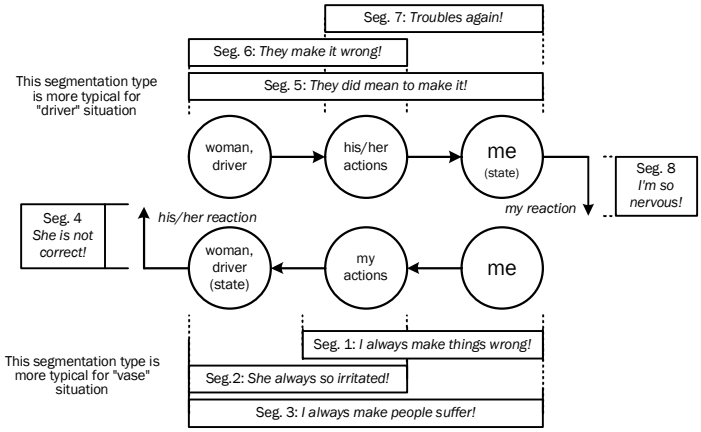

Fig. 1. Segmentation and extraction of a predicative structure from Rosenzweig situations

The scheme represents speaker ('me') making some actions or experiencing some state when being affected by some other actions. The actions of the speaker may affect the addressee – this interaction may be segmented in three different ways, depending on which components of the situation speaker takes into account: 'I make some actions' (Seg. 1), 'some actions (possibly not mine) affect the addressee' (Seg. 2) and a combination of the two – 'I make some actions affecting the addressee'

(Seg. 3). Further, the speaker may construct a representation for the addressee who has some true of false understanding of what has happened (Seg. 4), and the four segments should be symmetrically doubled to represent possible actions of the addressee in relation to the speaker ('myself') and speaker's understanding (Seg. 5-8). Each of these segments has a clear predicative structure, which can be processed in the model. During processing each segment may activate r-scripts with 'rational' speech output or d-scripts with 'emotional' reply – the latter is shown on the scheme for each segment.

In fact, most of the studied PFT situations may be represented by any of the selected segments, but there are still some preferences. 'Vase' situation is concentrated on the speaker's actions in respect to the addressee ('I broke his/her vase'), so the possible segmentation would be Seg.1-4 with the most complete – Seg. 3. On the other side 'driver' situation represents addressee's actions and is segmented by Seg. 5-8. Speaker may violate this tendency and treat 'vase' situation, for example, as Seg. 5 (*Is it a long time that the vase stands here?* – resulted by 'someone contrived it'), and 'driver' situation as Seg. 1 (*I have to be more attentive!* – resulted by 'I'm inadequate'). Each of the segments (represented as a predicative structure) arrives to the input of d/r-script set and activates one or several scripts with further speech output. A sample scheme for a segment processing is shown on Fig. 2.

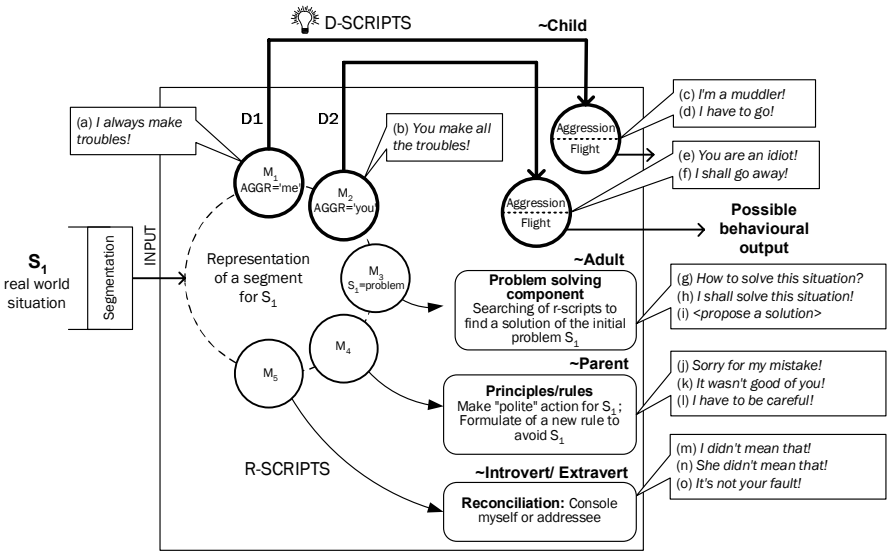

Fig. 2. Processing of a single component through d/r-scripts and possible outputs

D-script processing may differ depending on the actant of AGGR valency – it can be the speaker himself – D1 (*I always make troubles!*) or the addressee – D2 (*You make all the troubles!*). In all cases the speaker also occupies VICT valency, as he is suffering from the situation (other types of emotional communication like "communication of victims" and "speech persuasion", not observed here, may assign VICT valency to other participants). The selection between D1 and D2 is defined by the preferred segment: Seg. 1, 3, 8 activate a d-script in D1 scheme, and Seg. 2, 4, 5, 6

– activate D2 (we assume, that preference of reaction D1 or D2 influences extraction of a particular segment from the situation). The positions D1 and D2 may be occupied by different d-scripts from the 13 d-scripts list, e. g. we may treat other's actions (D2) ad inadequate, subjective, dangerous, etc. Each of the d-scripts may result speech output from it's starting models (M_1, M_2) with corresponding utterances falling in classes (a) and (b), and from it's final models of aggression or flight: classes (c-f).

The other way to process a segment is to activate an r-script. Since there was no given list of r-scripts or criteria for their selection (like critical elements for d-scripts), r-scripts were extracted from actual utterances basing on the similarity of their supposed initial models and semantics of speech output. Finally, selected r-scripts were categorized in three blocks: problem solving, principles/rules and reconciliation (minor speech output classes are here omitted). In problem solving block a speaker treats an initial segment as a problem and proposes a solution (i), offers to propose a solution (h) or requests a solution from the addressee (g). In principles/rules block a speaker follows etiquette (j), appeals to addressee to make him follow a rule (k) or formulates a new rule for himself or addressee (l). In reconciliation block speaker pays attention to his own emotional state or emotional state of the addressee (here emotional state is described as an activation of a d-script by speaker and/or addressee); here speaker has to detect a particular d-script (emotional state) and minimize it's activation by changing a representation of the initial situation. Speaker can represent his actions as less harmful for the addressee (m), justify addressee's actions for himself (n) or falsify the addressee's harmful actions (o).

The analysis in Fig. 2 has to be multiplied on the number of segments, selected in a particular situation with further combination and deletion of non-existing classes. For ‘vase’ situation with consideration of Seg. 1-5 this gives 28 general classes with the possibly of further classification (e.g. for emotional/reconciliation we may analyse, which exactly d-script is activated or is supposed to be activated etc.).

The segmentation allows to mark-up the utterances, received in the actual experimental study and to form a database. The proposed analysis offers more detailed classification than the standard PFT key (only 9 classes) and for many classes defines linguistic rules to construct utterances for this class in a speech interface. This allows not only to step from particular utterances to psychological characteristics (as in standard psychological interpretation of test results) but also to construct a set of possible utterances once a particular class is selected for output synthesis.

The next important task is to find out which class of utterances in d/r-script model is selected in a given situation for particular psychological characteristics. The question can be solved theoretically: we can assume, that aggressive participants may prefer D2 reaction while diffident participants – D1 reaction. We may suppose that a tendency to transfer control to a specific component of the model (as on Fig. 2) may correspond to different ego-states like “wish/fear-oriented” behaviour (d-scripts), “rational” behaviour (problem solving) or “rule-oriented” behaviour (principles/rules). On Fig. 2 we have proposed a shallow mark-up of the components, following E. Berne concept of ego-states: Child, Adult and Parent [2]. Further, reconciliation may be speaker-oriented (speaker calms himself) or addressee-oriented (speaker trys to calm the addressee). This may follow introvert / extravert psychological distinction. The other approach is to test the correspondence between psychological characteristics and preferred utterance experimentally.

4. Experimental study

The experimental study was carried out on several groups of respondents and included several PFT pictures (from 2 to 5) and supplementary tests, different for each group. We used an extended study of PFT, where the participants were asked several questions, in particular: *What would you say in this situation?* (WWYS) *What would you think?* (WWYT) and *What would you do, if you were not limited?* (WWYD). The following groups took part in the survey:

Adults – 100 respondents (2 PFT situations, no supplementary tests); University entrants – 110 respondents (5 PFT situations, Leary test, Lichko test); Practicing therapeutics and surgeons – 50 respondents (2 PFT situations, emotional degression test). The participants were native russian speakers, all the protocols were in Russian. The results of the survey were organized in a database, giving access to all actual utterances of a particular d/r-script class. The database can be used to adjust the representation of utterances in each specific class to meet actual speech practice. Further the database includes a set of psychological characteristics for participants.

Shallow results. The results for WWYS question for 'vase' situation remain quite stable and don't change for the three groups, staying at the following levels: (j) *Sorry! Excuse me!* – 43-46%, (h) *I shall compensate it!* – 26-30%, (m) *I didn't mean that!* – 14-21%. These three classes of results cover the majority of replies in WWYS (89-96%). This gives us a shallow understanding that a computer agent can be pre-programmed to produce a limited number of "etiquette" utterances and doesn't require any extensional reactions. The results however are motivated by the purpose of PFT – to verify the socialization of participants. Stability in replies do confirm the socialization, however, for an emotional agent we further require an adequate performance in WWYT and WWYD tasks, which can be usual for everyday communication and in addition may produce communicative stimulus, allowing the agent to initiate communication.

Answer filtering. For adults group the replies in WWYS and WWYT groups significantly differed. Some answers ((j) *Sorry*) are dedicated to the addressee – participants 'say' them, but don't 'think' this way, other classes ((a) *I'm a muddler!*) are selected in WWYT task and are suppressed in WWYS. We interpret the results as an activity of utterance filter, which rises the activation of "etiquette" classes (the answers are produced, even if there is a doubtful opportunity to prefer them), and lowers activation of some emotional classes, not allowed by politeness.

Answer filtering suggests to us a mechanism of class substitution: where an emotional utterance like (b) *You accuse me of what I haven't done!* is suppressed by the filter, it can exploit another more competent class like (m) *I didn't mean that!* giving to this utterance a distinctive emotional prosody (as in *Don't you see, that I didn't mean that?!*).

Correlations in Lichko test. Studies of personal psychological accentuations, measured in Lichko test, and classes of utterances, marked by d/r-scripts (University entrants, 5 PFT situations) give a list of correlations (relevant at $p<0,05$), in particular:

(i) psychasthenic level correlates with d-script answers: participants with psychasthenic accentuation prefer answers, defined by starting or target models of d-scripts – classes (a-f);

(ii) sensitivity – flight (d, f): sensitive participants tend to report flight plans, either because they are upset about their own actions (class (d)) or because they are affected by the actions of addressee (class (f)); further, if the addressee is responsible for the situation (e. g. he has spoiled out suit) sensitive participants don't justify his actions and don't propose answers (o) *It's not your fault!*

(iii) schizoid participants (as sensitive) don't justify the addressee in case of his fault;

(iv) level of sincerity correlates with d-script answers: sincere participants propose more "emotional" answers described by d-scripts, the answers may fall both in D1 (accuse myself) or D2 group (accuse the addressee);

(v) dissimulation correlates with extrovert d-script answers: insincere participants tend to propose emotional utterances, accusing the addressee – (b, e).

Correlations in Leary test. The results of Leary test give a list of correlations with the classes of produced utterances (relevant at $p<0,05$). Most of them seem trivial, but can be treated as a valuable dependency between "high-level" control states like emotional characteristics and "low-level" linguistic definitions of a particular utterance class, in particular: (vi) egoistic level negatively correlates with the sum of points for utterances, aimed at the addressee; (vii) aggressive level correlates with D2 (extroversive) d-script utterances (b, e, f).

Other results are less trivial and are worthy of notice, in particular: (viii) friendliness correlates with the sum of points for emotional utterances – on one side, and with the sum of D1 utterances plus sum r-script utterances, aimed at the speaker (like to consulate myself, formulate a rule for own actions etc.).

Correspondence between the number of emotional replies (assigned to d-scripts) and sincerity – on one side (Lichko test) and friendliness – on the other side (Leary test) may suggest a reverse dependency: to look friendly a computer agent might simulate emotions and report emotional utterances when interacting with a human in emotional situation. As expected, this behaviour must be refined by accurate application of critical elements to distinguish the case of sincerity from psychasthenic behaviour. As expected, "friendly" emotional interaction may contain both ironic aggression (Ironically: *Do you understand, what you've said?*) and sincere utterances (*Sometimes I make people suffer!*). These classes and their distinction from corresponding aggressive and self-accusation texts constitute one the most interesting areas for further studies.

5. Conclusion and Future Work

In the present report we have represented an application of a d-script model to mark-up of experimental data set and consecutive findings on the correlations between psychological characteristics of participants and classes of utterances, as defined by the theoretical linguistic model. As expected the correlations may allow creating a computer agent interacting with a human in a natural domestic or office environment and simulating certain mood or character in emotional speech interaction. As we propose, an important feature of the agent is to simulate and report it's internal emotional states thus setting up a friendly mood in communication.

Future work on the project is dedicated to the extension of psychological studies and extension of linguistic characteristics, assigned to each class of utterances. We intend: (a) to extend the created corpora of emotional utterances in Rosenzweig test for different age and professional groups of participants, linked with psychological characteristics – this allows to refine the structure for each class of utterances, as proposed by the d-script theory, and to produce accurate output in text synthesis tasks. (b) As the intonation serves as a distinctive feature for most of the classes – it is important to include it in the description; studies are based on the project "Intonation of a Russian dialogue" (Moscow State University). (c) PFT test considers situations where the dialogue is initiated by an event, in other cases it is important to detect and keep certain dialogue mood and sometimes to support a dialogue atmosphere, proposed by a speaker; as d-script model offers quite accurate means for definition of dialogue mood, the development of the corresponding classification is one of the future priorities. (d) Interjections offer a typical start for an emotional utterance but are actually omitted by participants of PFT; cooperating with project of Multimedia dictionary of interjections (RSUH) we work to equip the classes of emotional utterances in d-scripts model by typical interjections, to be produced during text synthesis.

References

1. Allen R. S. Concern Processing in Autonomous Agents. Ph. D. thesis. School of Computer Science, University of Birmingham (2001)
2. Berne, E. Games People Play: The Psychology of Human Relationships. Ballantine Books (1996)
3. Kotov, A. A. D-scripts model for speech influence and emotional dialogue simulation. In Proceedings of the 7th Annual Colloquium for the UK Special Interest Group for Computational Linguistics. University of Birmingham (2004) 134-140
4. Kotov, A. D-script model for synthesis and analysis of emotional speech : Proceedings of "Speech and Computer" SPECOM'2004, St.-Petersburg (2004) 579-585
5. Sloman, A. Beyond Shallow Models of Emotion. In Cognitive Processing, Vol. 1 (2001)
6. Sloman, A. Varieties of Affect and the CogAff Architecture Schema : C. Johnson (ed.), Proceedings Symposium on Emotion, Cognition and Affective Computing AISB'01 Convention, York, March (2001) 39-48
7. Sloman, A., Chrisley, R. Virtual Machines and Consciousness. In Journal of Consciousness Studies. Vol. 10, No. 4-5, April/May (2003) 133-172
8. Апресян В. Ю. Имплицитная агрессия в языке = Apresian U. Implicit aggression in speech. In Компьютерная лингвистика и интеллектуальные технологии: Тр. Междунар. конференции Диалог 2003. – М.: Наука (2003) 32-35
9. Гловинская М. Я. Гипербола как проявление и оправдание речевой агрессии = Glovinskaya M. Hyperbole as an expression and justification of speech aggression. In Сокровенные смыслы. Сборник статей в честь Н. Д. Арутюновой (2004) 69-76
10. Теория и практика лингвистического анализа текстов СМИ в судебных экспертизах и информационных спорах = Theory and practice of linguistic analysis of mass media texts in juridical expertises and information disputes. In Сборник материалов научно-практического семинара. Москва 7-8 декабря 2002 г. Часть 2. – М.: Галерия (2003)