

Michael Kipp. ANVIL 4.5. Annotation of Video and Spoken Language. <http://www.dfki.de/~kipp/anvil>

Isabel Galhano Rodrigues

*Faculdade de Letras da Universidade do Porto (Portugal)*

O programa ANVIL é uma ferramenta desenvolvida por Michael Kipp<sup>1</sup> (cf. Kipp 2004) para a análise de material audiovisual digitalizado. Pela sua simplicidade, pode dizer-se que é um instrumento eficaz e preciso que veio substituir o trabalho moroso e, de certo modo, subjectivo, de anotações por meio de observação e bloco de notas. Segundo o autor, o objectivo principal de Anvil é precisamente o de facilitar o processo de transcrição, tornando a codificação das unidades verbais e não-verbais o mais intuitivo e fácil possível. Na verdade, ANVIL possibilita um registo de dados preciso, sobretudo no que diz respeito a questões de alinhamento da fala e dos movimentos do corpo, assim como da coordenação entre as várias modalidades de movimento<sup>2</sup>. Tem grande proveito nas áreas científicas da Antropologia, Psicologia, Sociologia, Linguística, Etologia e até mesmo da Inteligência Artificial.

Sob o ponto de vista técnico, ANVIL apoia-se na linguagem de programação *Java* e na linguagem de estruturação XML. Por isso, o programa corre numa plataforma *Java* em Windows, Linux, Solaris ou Macintosh (neste último mais difícil de instalar). Como os ficheiros

---

<sup>1</sup> "Anvil was written as part of a PhD project on nonverbal communication at the University of Saarland [...] on a grant by the DFG (Deutsche Forschungsgemeinschaft) and builds on experiences with mass corpus annotation of dialogue acts within the speech-to-speech machine translation project VERBMOBIL" (Kipp 2003: 6).

<sup>2</sup> Por modalidades de movimento entende-se os movimentos feitos por diversas partes do corpo, por exemplo, os movimentos dos olhos, do tronco ou da cabeça.

ANVIL se baseiam em XML, qualquer utente os pode manipular, de acordo com os seus interesses, recorrendo a ferramentas de transformação à disposição em quase todos os *browser* da Internet.

Com ANVIL não é possível fazer transcrições fonéticas ou prosódicas da fala. O autor sugere que estas sejam feitas previamente com o programa PRAAT<sup>3</sup>. Este programa permite criar ficheiros de texto contendo as transcrições que é depois possível importar para um projecto em ANVIL.

ANVIL também não permite fazer qualquer tipo de análise estatística dos dados anotados. Para isso, o autor aconselha a que se recorra a outros produtos comerciais de *software*, como por exemplo SPSS ou Statistica (Kipp 2003: 6).

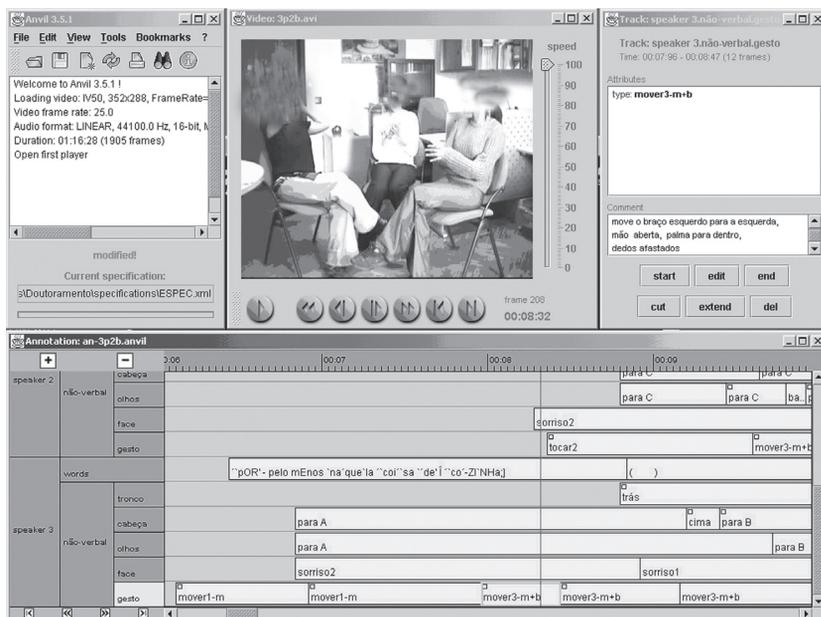
No manual (*user manual*) que se pode obter na página de *download* do programa (Kipp 2003) encontram-se as instruções detalhadas e bem estruturadas sobre:

- o modo como fazer a instalação em Windows. Qualquer utente, mesmo com poucos conhecimentos de informática, não terá dificuldades em instalar o programa. Em último caso, poderá sempre recorrer a um apoio personalizado *on-line*.
- os ficheiros vídeo reconhecíveis por ANVIL e ao modo de os fazer.
- o conceito que está por trás do programa, assim como a sua estrutura hierárquica, organização de fiadas e de ficheiros.
- a interface do utente (através de ilustrações e explicações simples e detalhadas): a janela principal, a janela vídeo, a janela dos elementos e o quadro de anotação.
- o trabalho de anotação usando o quadro de anotação.
- importação da transcrição feita com PRAAT.

---

<sup>3</sup> Programa desenvolvido por Boersma e Weenink, Universidade de Amsterdão, 1992-1998. Página de *download*: [http://www.fon.hum.uva.nl/praat/download\\_win.html](http://www.fon.hum.uva.nl/praat/download_win.html)

- descrição da elaboração de ficheiros áudio e da transcrição com PRAAT.
- trabalho com projectos e exportação de fiadas.
- elaboração de especificações em XML, com exemplos



MARCAÇÃO DA VARIÁVEL GESTO DA FALANTE 3 (AT); COMENTÁRIO SOBRE PORMENORES NO QUADRO SUPERIOR DIREITO.

ANVIL permite observar um ficheiro vídeo em câmara lenta e com som. O filme pode ser aberto sozinho, sem mais quadros no interface.

Para anotar os movimentos do corpo no vídeo é preciso abrir todas as janelas e o quadro de anotação. Os items que se encontram no quadro e a que correspondem as fiadas onde se podem etiquetar os movimentos são estruturados em XML, numa especificação. Podem elaborar-se diversas especificações conforme as variáveis que se pretenderem analisar. A cada item no quadro (atributos) podem fazer-se corresponder várias variáveis (valores).

Para uma das fiadas do quadro de anotação importa-se a transcrição da fala, previamente realizada com o programa PRAAT. A transcrição terá de ser ortográfica, podendo conter sinais correntes na transcrição prosódica, como por exemplo  $\hat{\ } , ' ; .$

Numa outra fiada pode também introduzir-se o ficheiro com as informações correspondentes à altura de tom ou à amplitude (igualmente importado de PRAAT). As restantes fiadas estão destinadas à etiquetagem dos valores estruturados e correspondem ao que está a ser dito e feito no filme. Por exemplo, para a fiada correspondente aos movimentos da cabeça, podem considerar-se os valores *para cima*, *para baixo*, *para o lado direito* ou *para o lado esquerdo*. Além destes valores, na janela dos elementos podem escrever-se comentários relativos ao movimento que parece necessário focar (por exemplo: *movimento abrupto, simultâneo à verbalização da palavra "quê", com uma subida de altura de tom e aumento da intensidade da voz*). Veja-se também, na imagem a seguir, o comentário escrito na janela em cima, à direita, para o gesto etiquetado com a expressão *mover3-b+m* (mover braços e mãos).

O quadro de anotação está sincronizado com o filme: a linha vermelha do cursor, perpendicular ao quadro de anotação, assinala o momento do filme correspondente à imagem na janela vídeo. Assim, este cursor abrange todas as fiadas e mostra o que deve registar-se, ou o que já se encontra registado para cada momento relativamente ao que é dito e feito.

A linha do tempo no quadro de anotação pode ainda ser encurtada ou aumentada conforme as necessidades de cada um. No caso de uma micro-análise, cada traço na linha do tempo, corresponde a uma imagem do filme.

A possibilidade de usar mais do que um quadro de anotação (ou seja, diversas especificações) por ficheiro vídeo torna a aplicação do programa muito abrangente e flexível. Imagine-se que, numa primeira fase, se procede apenas à descrição dos movimentos usando uma especificação e, numa segunda, à sua classificação

com outra especificação. Não quer dizer que não se possa introduzir uma especificação estruturando fiadas só destinadas à descrição do movimento e fiadas destinadas à classificação desses mesmos movimentos. As limitações impostas são simplesmente causadas por questões de *hardware*, ou seja, pelas medidas do écran. Demasiadas fiadas de uma só vez dificultam a análise, pois, para interpretar o que foi anotado em cada fiada, torna-se necessário mover constantemente o quadro de anotação para cima e para baixo. O mais fácil é recorrer a várias especificações conforme os objectivos de cada etapa da análise. Também é possível trabalhar com dois computadores, e abrir em cada o mesmo ficheiro vídeo com diferentes anotações.

Uma das desvantagens deste programa é o facto de os ficheiros vídeo não deverem ultrapassar os 40 MB, o que significa que não podem ter mais do que 2 minutos (conforme a conversão do vídeo e os codecs escolhidos). Outra prende-se com o quadro de anotação. Na maior parte dos computadores em que se instala ANVIL, não é possível imprimir a anotação dos movimentos. O autor espera que este problema possa vir a ser resolvido. Por enquanto, sugere que se façam impressões da imagem do écran (*print screen*), no caso de se pretender um apoio em papel de algum aspecto mais importante da anotação (a imagem atrás foi conseguida através deste processo).

Uma grande vantagem deste programa é ainda o facto de o autor estar à disposição para dar apoio no uso do programa, procurando soluções para resolver os problemas particulares de cada um.

Como linguista que investiga as modalidades verbais e não-verbais da interacção, procurando dar conta, num trabalho intensivo de micro-análise, da sequência dos enunciados, da prosódia, da proxémica, dos movimentos da cabeça, do tronco, dos braços e mãos, da face e dos olhos, encontrei no ANVIL uma ferramenta fácil, intuitiva e muito versátil (cf. Rodrigues 2003).

**REFERÊNCIAS**

- Kipp, M. 2003. *Anvil 4.0, Annotation of video and spoken language. User Manual*. <http://www.dfki.de/~kipp/anvil/doc/anvil40-manual.pdf> , acedido em 13.03.2006.
- Kipp, M. 2004. *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Boca Raton, Florida: Dissertation.com.
- Rodrigues, I. M. G. 2003. *Fala e movimentos do corpo na interação face a face. Estratégias de reparação e de (des)focalização na manutenção de vez. Tese de Doutoramento*. Porto: Faculdade de Letras da Universidade do Porto [no prelo].