Documentation for the Program NEWSPELL (Version 2.0)

by Rainer Pischner

1 General Information

Disclaimer: The program "NEWSPELL" is provided on an "as is" basis, without warranty of any kind.

NEWSPELL is a program that allows the manipulation of Spell files. One can use it for any of the following:

- Create distinctive calendars and eliminate multiple cites with a user defined hierarchy. For example, in the Spell file ARTKALEN.DAT (Occupation calendar on monthly basis) one can define a priority of spell information in case of multiple cites in such a way, that the event "unemployment" is picked over any other response given for the same time period.
- Aggregation of events. Different events can be combined within a single Spell type. For example the occupation status "Full-time Employment" and "Part-time Employment" can be combined into a single Spell type "Gainful Employment".
- Disaggregation of combined events. For example the events "Full-time Employment" and "Unemployment" can be split into three different categories, depending on their occurrence in a given time period of interest (month or year):
    - Type 1: Unemployed and full-time employed at the same time
    - Type 2: Unemployed but not full-time employed at the same time
    - Type 3: Full-time employed but not month unemployed at the same time
- The output can be selected as a new Spell file with censor information and/or as time series data.
- Additionally, a log-file keeps record of all activities of a session with NEWSPELL.
- ***NEW!!!*** A result file contains frequencies of each new spell type for each person.

NEWSPELL supports all GSOEP-Spell files, normally included in the yearly data distribution. These are:

- PBIOSPE.DAT Combined activity calendars (by years of age, starting at age 15)
- ARTKALEN.DAT Activity calendars (by month, starting in January 1983=1)
- EINKALEN.DAT Income calendar (by month, starting in January 1983=1)
- BIOMARSY.DAT Marital status (by years of age, starting at age 0)
- BIOMARSM.DAT Marital status (by month, starting in January 1983=1)
- SOZKALEN.DAT Social Assistance received by household (by month, starting in Jan. 1991)

Note:	When working with GSOEP spell files, the user needs to give the password of the GSOEP data distribution that a given version of NEWSPELL came with!

Spell files created by the user can be processed with NEWSPELL as long as certain rules are followed. Since the structure of spell-oriented output files produced by NEWSPELL is very similar to the above mentioned spell files, they can again be used as input for further processing, if necessary.

The program NEWSPELL is supplied with a command file (as in RZOO/TDA) when run.
The command to execute NEWSPELL is:

NEWSPELL[.EXE] command file [password]

As previously mentioned the password needs to be given when using GSOEP Spell files, only.
Example:

NEWSPELL MY.CMD or **NEWSPELL ARTKALEN.CMD secret**

"Installation" of NEWSPELL

The best and easiest way to "install" NEWSPELL is to copy all files from the folder NEWSPELL on the GSOEP CD-ROM to any work directory on your hard disk.

2 The command file (.cmd)*

The command file consists of a sequence of commands which almost always have the same structure:

Type of command = command value

Note:	The input commands are not case sensitive, i.e. capital letters are treated the same way as small letters.

The following section describes the available commands in detail.

2.1 Comment lines and comments embedded in command lines

Any line starting with the symbols /* is interpreted as a comment line and not executed. Within a command line anything to the right of /* is also ignored by the program.

/* this is a comment

2.2 Defining the path of the input file

This command sets the path where the input file is located. The directory must have been created before running the program and write permission to this path must be given.

PI = Path of Input file (e.g. PI = C:\NEWSPELL\ )

If the command PI is not found then it is assumed that NEWSPELL was started using an input file from the current directory (or folder). The following is therefore the default

PI =Current directory (i.e. PI = :)

2.3 Defining the input file

This command tells the program the name of the input Spell file.

NI = Name of Input file

Example:

NI = EINKALEN.DAT

The default Spell file is NI = ARTKALEN.DAT and is used when the command NI is omitted.

2.4 Defining the path for the output file

PO = Path of Output-file

This command tells the program where to store the resulting files. The directory must have been created before running the program and write permission to this path must be given.
The default is

PO = PI (= Path of input file)

If one would like to use the current directory, i.e. the directory where the program NEWSPELL is started from, then the command can be omitted:

2.5 The output files

2.5.1 The log-file

NL = Name of log-file (protocol of session)

This command sets the name of the protocol-file which keeps record of activities of the current session. The default is NL = NEWSPELL.LOG

2.5.2 The result-file ***(NEW!!! in 2004)***

NR = Name of result-file (some frequencies)

This command sets the name of the result-file with frequencies of new defined spell types for each person of the current session. The default is NR = NEWSPELL.RES

2.5.3 File with new defined time series data

NT = Name of output-file (Time-series)

This command sets the name of the output file, which will contain the resulting information as time series data. Example:

NT = EINKALEN.TIM /* Income calendar information per month

The default is NT = NEWSPELL.TIM

Variables in this output file include:

NEW!!!

HHNR	Fixed ID of wave 1 household
PERSNR/HHNRAKT	Fixed individual ID or current household-ID
BEGIN	Begin of Time-Series ***NEW: Point in time of the first valid observation***
END	End of Time-Series ***NEW: Point in time of the last valid observation***
SPELLTYP(BEGIN)	New spell type at time BEGIN
SPELLTYP(BEGIN+1)	New spell type at BEGIN+1
.....	
SPELLTYP(END)	New Spell type at time END

The variables in the output file are separated by one or more blanks.

The command

NT = none

suppresses the output of this file.

2.5.4 File with new defined Spell information

NS=Name of output-file (Spells)

This command sets the name of the output file which will contain the resulting information as spell data.
Example:

NS=EINKALEN.SPL /* Income calendars as Spells

The default is NS=NEWSPELL.SPL

Variables in this output-file include:

HHNR	Fixed ID of wave 1 household
PERSNR / HHNRAKT	Fixed individual ID or current household-ID
SPELLNR	Serial Spell number
SPELLTYP	New Spell type
BEGIN	New begin of Spell
END	New end of Spell
SPELLTYP(BEGIN-1)	Previous Spell type ....
SPELLTYP(END+1)	Following Spell type
ZENSOR	Censor with -2 = Spell type is not defined 1 = Uncensored 2 = Normal right censor = N-RZ 4 = Normal left censor = N-LZ 5 = Right and left censored N-RZ, N-LZ

Definition of ZENSOR

Assuming:
1. *TB is defined as the begin of OLDSPELL, and tb as the begin of NEWSPELL*
2. *TE is defined as the end of OLDSPELL, and te as the end of NEWSPELL*
3. *The transition is defined as*

NEWSPELLTYPE(t) is unequal to NEWSPELLTYPE(t+1) , where t,t+1 lie in the open interval (tb,te)

*Then we have a **left-censored** spell,*
if NEWSPELLTYPE(t) has not a valid value or
t is equal to TB or
the begin of NEWSPELLTYPE(tb) is equal to NEWSPELLTYPE(tb-1)

*Similar there is a **right-censored** spell, if*
SPELLTYPE(t+1) has not a valid value or
t+1 is equal to TE
or the end of NEWSPELLTYPE(te) is equal to NEWSPELLTYPE(te+1)
Special case: if NEWSPELLTYPE has not a valid value the corresponding censor get the value -2 (TNZ).

The variables are separated by one or more blanks.

This command

NS=none

suppresses the output of the file.

2.6 Defining time period of interest (BEGIN and END)

Up to wave 14 (1997), the files ARTKALEN, EINKALEN, and BIOMARSM include spell information on up to 168 months. PBIOSPE and BIOMARSM cover a given respondent's life span, starting with age 15 and ending with the age reached at the point of time when the most recent interview was carried out. NEWSPELL allows the researcher to restrict the time period for any given analysis by defining a new begin and end information according to the research question.

With the commands

NB=New Begin (e.g. NB = 73 [January 1989])

and

NE=New End (e.g. NE = 144 [December 1994])

selected target areas can be chosen.

Note:	Using these commands in most cases reduces not only the number of spells per observation (individual or household) but the number of spell systems as well.

2.7 Manipulation of the Spell types

The commands for working with Spells are kept simple: They are only made up of a list of Spell definitions. The first listed definition has the highest priority, then the next definition receives the next highest priority, etc. Each command line begins with the definition of a new Spell type number, followed by an equal sign (=); the right hand side is given by a list of one or more original SPELLTYP values. Theses are linked with the @ symbol. For @ one can either substitute the logical "or" or "and". Also each line can have an optional comment which starts with the symbol /*.

Note:	In version 1.2 of NEWSPELL only the numbers 1-14 ,-1 (Missing-Value) and 99 (gap) are allowed as values of SPELLTYP. One could define other Spell numbers but then NEWSPELL cannot be used for any additional processing!

The Spell definition syntax is:

New_Spelltype_1=Old_Spelltype_x1 [@ Old _Spelltype_x2[@ Old _Spelltype_x3...]]]
New_Spelltype_2=Old_Spelltype_y1 [@ Old _Spelltype_y2[@ Old _Spelltype_y3...]]]
...
New_Spelltype_3=Old_Spelltype_z1 [@ Old _Spelltype_z2[@ Old _Spelltype_z3...]]]

with @ = or / and . The information in brackets is optional.

Note:	The priority of a Spell definition is set according to the order in which the manipulation command lines are listed and not by the values of the new Spell information.

Examples:

The original values of SPELLTYP are defined as follows:

1 = Education/training
2 = Full-time employed
3 = Part-time employed
4 = Unemployed
5 = Retired
6 = Other

Example 1: (Hierarchy and Aggregation)

Assuming, we are analyzing "Unemployment", "Employment", and "Other" with a primary focus on "Unemployment".

The following command lines give "Unemployment" priority over every other activity:

```
1=4          /* Unemployed                        =>1st Priority
2=2 or 3     /* Full-time and part-time employed   =>2nd Priority
3=1 or 5 or 6 /* Other                           =>3rd Priority
```

The following alternative command lines can also be used but it is not recommended because NEWSPELL only allows a limited number of process statements (maximum 14).

```
1=4          /* Unemployed                        => 1st Priority
2=2          /* Full-time employed                => 2nd Priority
2=3          /* Part-time employed                => also 2nd Priority
3=1          /* Education/training                => 3rd Priority
3=5          /* Retired                           => also 3rd Priority
3=6          /* Other                             => also 3rd Priority
```

Example 2 (Hierarchy, Aggregation, und Disaggregation):

Assuming, we are interested in the simultaneous occurrence of , "Unemployment" and "Employment" as first priority, and the unique occurrence of each of these labor market status is less important. This can be achieved by using the commands:

```
1=4 and 3 or 4 and 2
             /*  Unemployed and full-time employed
             /*  Or unemployed and part-time employed   =>1st Priority
2=4          /*  Unemployed                             => 2nd Priority
3=2 or 3     /*  Full-time employed or part-time employed =>3nd Priority
4=1 or 5 or 6 /* Other                                  =>4rd Priority
```

The alternative - not recommended - is:

```
1=4 and 3    /* Unemployed and full-time employed       => 1st Priority
1=4 and 2    /* Unemployed and part-time employed because
             /* New spell type is identical with that in the previous line.  => also 1st Priority
2=4          /* Unemployed                             => 2nd Priority
3=2          /* Full-time employed                     => 3nd Priority
3=3          /* Part-time employed                     => also 3nd Priority
4=1          /* Education/training                     => 4rd Priority
4=5          /* Retired                                => also 4rd Priority
4=6          /* Other                                  => also 4rd Priority
```

Note: If definition is missing in a new spell type then only those Spells are produced where the Spell types with a lower number have a higher priority. In the example given above Spell type 1 (education/training) would always win out over other activities at that point in time.

2.8 Gaps ***NEW!!!***

The handling of gaps is changed. Three types of gap are calculated:

The formerly use code –1 for gaps is no longer used. It is grouped in three types of gaps:

Code 97 marks a gap before the appearance of the first valid activity,
Code 98 marks a gap after the appearance of the last valid activity,
Code 99 marks a gap between valid activities.

The censor variable of gap spells get the code –2 (not applied)

2.9 User Defined Spells

Besides the standard data sets contained in the SOEP one can also work with user defined spell files and those, which have previously been created by NEWSPELL. In any case, the following data structure is required:

CASE-ID <D> PERSNR /HHNRAKT <D> SPELLNR <D> SPELLTYP <D> SPELLBEGIN <D> SPELLEND

with <D> as a delimiter . You can choose as delimiter blanks, tabs, comma or semicolon

Note: Differently to earlier versions of NEWSPELL you do not need the command FM = delimiter

3 Concluding comments

It is highly recommended that the user stays with this syntax because of the limited error checking. The users should also make sure that the results are as expected and correct!

4 Examples for Command files (*.CMD)

4.1 Artkalen: creating unique spell information without re-definition of original spell data

```
PI=:              /* sub-directory containing input file
NI=ARTKALEN.DAT   /* Input-File
PO=:              /* sub-directory for output
NT=ARTKAL1.TIM    /* output-file (time series data)
NS=ARTKAL1.SPL    /* output-file (spell data)
NL=ARTKAL1.LOG    /* Log-file

/* no further specification or re-definition of spelltypes
/* default output is unique spell-information with ...
/* Spell type = 1 is  1. priority
/* Spell type = 2 is  2. priority  etc.

/* time period of interest: Wave 13 (Jan-Dec. 1995)
nb=145
ne=156
```

4.2 Artkalen: differentiating 3 types of monthly labor market status

```
PI=:               /* subdir input
NI=ARTKALEN.DAT    /* input-file
PO=:               /* subdir output
NT=ARTKAL2.TIM     /* output-file  (time series data)
NS=ARTKAL2.SPL     /* output-file  (spell data)
NL=ARTKAL2.LOG     /* Log-file

/* New Spell differentiates 3 types of employment status
1=1 or 2 or 3 or 4 or 11        /* Employed
2=5                             /* Registered Unemployed
3=6 or 7 or 8 or 9 or 10 or 12  /* Not employed

/* time period of interest: Waves 7-12 (Jan. 1989 - Dec. 1994)
NB=73              /* Begin 7. Wave
NE=144             /* End  12. Wave
```

4.3 Artkalen: differentiating simultaneous spell information on monthly labor market status

```
PI=:               /* sub-directory containing input file
NI=ARTKALEN.DAT    /* Input-File
PO=:               /* sub-directory for output
NT=ARTKAL3.TIM     /* output-file (time series data)
NS=ARTKAL3.SPL     /* output-file (spell data)
NL=ARTKAL3.LOG     /* Log-file

/* Analysis interest:
/*   1. Unemployment PLUS any type of employment
/*   2. Unemployment ONLY
/*   3. Employment ONLY
/*   4. Other

/* definition of new spell type
1=5 and 1 or 5 and 2 or 5 and 3 or 5 and 4 or 5 and 11
2=5
3=1 or 2 or 3 or 4 or 11
4=6 or 7 or 8 or 9 or 10 or 12

/* time period of interest: Waves 10-11 (Jan. 1992 - Dec.
1993)
nb=109
ne=132
```

4.4 PBIOSPE: yearly labor market status of persons aged 50-65 years

```
PI=:
NI=PBIOSPE.dat
```

```
PO=:
NT=pbio4.tim    /* output time-series data
NS=pbio4.SPL    /* output SPELL-data
NL=pbio4.LOG    /* Logfile

/* differentiating labor market status
/*         1.   education and training
/*         2.   employment
/*         3.   non-employment
/*         4.   pensioner
1=1 or 2
2=4 or 5
3=3 or 6 or 7 or 9
4=8

/* time period of interest: age 50-65
nb=50
ne=65
```

4.5 BIOMARSY: yearly marital status of persons aged 20-30 years

```
PI=:              /* subdir input
NI=BIOMARSY.DAT   /* input-file
PO=:              /* subdir output
NT=MARSY5.TIM    /* output-file  (time series data)
NS=MARSY5.SPL    /* output-file  (spell data)
NL=MARSY5.LOG    /* Log-file

/* New Spell differentiates 3 types of marital status
1=1              /* Single
2=2              /* married
3=3 or 4 or 5    /* No longer married

/* time period of interest: Age 20-30
NB=20
NE=30
```

4.6 Using self-defined spell data

```
PI=:
PO=:
NI=DEMO.DAT         /* Input Spell-File     DEMO
NT=EXAMDEMO.TIM        /* output-file (time series data)
NS=EXAMDEMO.SPL        /* output-file (spell data)
NL=EXAMDEMO.LOG        /* Log-file

/* new spell with code 1,
    if old spell was coded 1,2,3,4,11
/* new spell with code 2,
    if old spell was coded  5,6,7,8,9,10,12
1=1 or 2 or 3 or 4 or 11
2=5 or 6 or 7 or 8 or 9 or 10 or 12
```